

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:36:00 ; Search time 4.06349 Seconds
(without alignments)
3105.882 Million cell updates/sec

Title: US-09-813-214a-11
Perfect score: 196
Sequence: 1 GIGISEADGKGANARGDKSTAIGDIAQALGSQSTIAIGD 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	66.3	1964	2 Q8KQM9	Q8kqm9 moraxella c
2	84	42.9	2314	2 Q8KQM8	Q8kqm8 moraxella c
3	77.5	39.5	2190	16 Q8PGS0	Q8pgs0 xanthomonas
4	77	39.3	2351	16 Q8PCQ5	Q8pcq5 xanthomonas
5	71.5	36.5	1328	2 Q9LAX0	Q9lax0 xanthomonas
6	71	36.2	1299	16 Q9FX36	Q9fx36 pasteurella
7	67.5	34.4	382	16 Q8XJ23	Q8xj23 shigella fl
8	67.5	34.4	1588	16 Q8XDG4	Q8xdg4 escherichia
9	67	34.2	762	16 Q89IU6	Q89iu6 bradyrhizob
10	67	34.2	1265	2 Q9FDA0	Q9fda0 xanthomonas
11	67	34.2	1291	16 Q92KQ7	Q92kq7 rhizobium m
12	67	34.2	1778	16 Q8FCB2	Q8fcb2 escherichia
13	66	33.7	261	11 Q9EQT3	Q9eqt3 mus musculus
14	66	33.7	1004	16 Q9PD63	Q9pd63 xylella fas
15	66	33.7	2712	16 Q9FX35	Q9fx35 pasteurella
16	65.5	33.4	1315	16 Q87D62	Q87d62 xylella fas

17	65	33.2	487	2	Q9LA53	Q9la53 escherichia
18	65	33.2	1461	16	Q8ZL64	Q8zl64 salmonella
19	65	33.2	1953	16	Q98HJ2	Q98hj2 rhizobium l
20	64.5	32.9	1114	16	Q8PGR8	Q8pgr8 xanthomonas
21	64.5	32.9	1190	16	Q8PC04	Q8pc04 xylella fas
22	63	32.1	615	16	Q05806	Q05806 mycobacteri
23	63	32.1	615	16	Q7TXQ0	Q7txq0 mycobacteri
24	63	32.1	663	16	Q8VJCO	Q8vjco mycobacteri
25	63	32.1	694	16	Q53212	Q53212 mycobacteri
26	63	32.1	997	16	Q87DF4	Q87df4 xylella fas
27	63	32.1	1264	2	Q8RO61	Q8rq61 actinobacil
28	62	31.6	980	16	Q8A3J1	Q8a3j1 bacteroides
29	62	31.6	1309	16	Q8XPL8	Q8xpl8 ralstonia s
30	61.5	31.4	148	10	Q8S608	Q8s608 oryza sativ
31	61.5	31.4	2342	5	O01677	O01677 bombyx mori
32	61	31.1	369	16	Q89J38	Q89j38 bradyrhizob
33	61	31.1	454	2	Q8S267	Q8s267 yersinia en
34	61	31.1	511	9	Q9MCI8	Q9mc18 bacterioph
35	61	31.1	641	16	Q8CKM1	Q8ckm1 yersinia pe
36	61	31.1	658	16	Q8ZHQ0	Q8zhj0 yersinia pe
37	61	31.1	736	16	Q8CL86	Q8cl86 yersinia pe
38	61	31.1	3165	16	Q8XYI3	Q8xyi3 ralstonia s
39	60.5	30.9	1711	10	Q7XRT3	Q7xtt3 oryza sativ
40	60.5	30.9	2059	16	Q9PD50	Q9pd50 xylella fas
41	60	30.6	274	5	Q95RJ4	Q95rj4 drosophila
42	60	30.6	286	10	Q94CM0	Q94cm0 arabidopsis
43	60	30.6	286	10	Q8GYJ3	Q8gyj3 arabidopsis
44	60	30.6	289	10	O48846	O48846 arabidopsis
45	60	30.6	449	5	Q9V7C0	Q9v7c0 drosophila

ALIGNMENTS

RESULT 1

Q8KQM9 ID Q8KQM9 PRELIMINARY; PRT; 1964 AA.
AC Q8KQM9;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAG.
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O35E;
RX MEDLINE=22112901; PubMed=12117964;
RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Geme J.W. III,
RA Hansen E.J.;
RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
RT Activities.";
RL Infect. Immun. 70:4523-4533 (2002).
DR EMBL; AY077637; AAL78284.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 9.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 1964 AA; 201585 MW; 764A08F5F1F6854E CRC64;

Query Match 66.3%; Score 130; DB 2; Length 1964;

Best Local Similarity 70.0%; Pred. No. 1.2e-06;

Matches 28; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 GIGISEADGKGANARGDKSTAIGDIAQALGSQSTIAIGD 40

Db 68 GIAIGSGGLNGGARADGKSIAGSTAQATGSGSIAIGD 107

```

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB012003; AAM38389.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 53.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2190 AA; 204715 MW; 1BEA011869A3C2CA CRC64;

Query Match 39.5%; Score 77.5; DB 16; Length 2190;
Best Local Similarity 45.2%; Pred. No. 2.5;
Matches 19; Conservative 7; Mismatches 11; Indels 5; Gaps 1;

OY 3 GISEADGGK-----GGANARGDKSIAIGDIAQAALGQSQSIAG 39
Db 1048 GSEFAGAGQSTALGAAGAYGDSLVAGLSQAQGSSESTAMG 1089

RESULT 4
Q8PGQ5 PRELIMINARY; PRT; 2351 AA.
AC Q8PGQ5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XCC0658.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cicarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB012164; AAM39974.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 59.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2351 AA; 220261 MW; C5482A38C940DAll CRC64;

Query Match 39.3%; Score 77; DB 16; Length 2351;
Best Local Similarity 60.7%; Pred. No. 3.2;
Matches 17; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 12 GGNARGDKSIAGIDIAQALGSQSIAG 39
Db 863 GOSNAAGDESIALGWEAQAEGDQIALG 890

RESULT 5
Q9LAX0 PRELIMINARY; PRT; 1328 AA.
AC Q9LAX0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Xanthomonas campestris pv. pelargonii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=91612;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xpel-1.
RA Kim J.F., Zumoff C.H., Beer S.V.;
RT "An alanine-, glycine-, and serine-rich protein and a putative serine
RT protease of Xanthomonas campestris pv. pelargonii.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083618; AAF63394.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 26.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1328 AA; 127402 MW; ACE01A4C6548364C CRC64;

Query Match 36.5%; Score 71.5; DB 2; Length 1328;
Best Local Similarity 46.5%; Pred. No. 7.7;
Matches 20; Conservative 5; Mismatches 13; Indels 5; Gaps 2;

QY 3 GISEADG---GKGG--ANARGDKSIAGIDIAQALGSQSIAGD 40
Db 236 GLSTAGSLSSAGGYLSRASGDASTAFGYRARGSSSIAGVD 278

RESULT 6
Q9F3X6 PRELIMINARY; PRT; 1299 AA.
ID Q9F3X6;
AC Q9F3X6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MapA protein (Hsf).
GN MAPA OR HSF.2 OR PM1570.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AJ277635; CAC14202.1; -.
DR EMBL; AE006194; AAK03654.1; -.
DR InterPro; IPR008640; Hep_Hag.

```

```

DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 10.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB6CDB428 CRC64;

Query Match 36.2%; Score 71; DB 16; Length 1299;
Best Local Similarity 45.5%; Pred. No. 8.7;
Matches 20; Conservative 7; Mismatches 11; Indels 6; Gaps 2;

QY 2 IGISEADGKGG---GANARGDK--SIAIGDIAQALGSQSIAG 39
Db 109 IGFGATNDGETNVAIGAKSKSKAASIAIGDNAKALDQAIAIG 152

RESULT 7
Q83J23 PRELIMINARY; PRT; 382 AA.
ID Q83J23;
AC Q83J23;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Orf, hypothetical protein.
GN SF3640.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; AE015372; AAN45087.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 382 AA; 37760 MW; F686328578741C10 CRC64;

Query Match 34.4%; Score 67.5; DB 16; Length 382;
Best Local Similarity 33.3%; Pred. No. 6.1;
Matches 18; Conservative 8; Mismatches 9; Indels 19; Gaps 2;

QY 5 SEADG---GKGG-----GANARGDKSIAGIDIAQALGSQSIAG 39
Db 267 NKADGVDAIALGNGSQSRGLNTALGTASNTGDKSLALGSNSANGINSVALG 320

RESULT 8
Q8XDGA PRELIMINARY; PRT; 1588 AA.
ID Q8XDGA;
AC Q8XDGA;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin (Hypothetical protein).
GN Z5029 OR ECS4480 OR H161.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

```

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.,
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki G., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 0509952;
 RA Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,
 RA Shinagawa H.,
 RT "O157 specific gene similar to H. influenzae adhesin gene.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005587; AAG58749.1; -;
 DR EMBL; AP002566; BAB37903.1; -;
 DR EMBL; AB036416; BAB87814.1; -;
 DR PIR; AB6036; AB6036;
 DR PIR; H91188; H91188;
 DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep_Hag; 13.
 DR Pfam; PF05662; HIM; 12.
 DR Pfam; PF03895; Yada; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1588 AA; 160150 MW; B2BA4E06BFF28DEC CRC64;
 Query Match 34.4%; Score 67.5; DB 16; Length 1588;
 Best Local Similarity 33.3%; Pred. No. 28;
 Matches 18; Conservative 8; Mismatches 9; Indels 19; Gaps 2;
 QY 5 SEADG-----GKG-----GANARGDKSIAIGDIAQAALGSQSIAG 39
 DB 281 NKADGVDAIALGNGSQSRGLNTIALGTASNATGDKSLALGSSNGSINSLVAG 334
 RESULT 9
 Q89IU6 PRELIMINARY; PRT; 762 AA.
 AC Q89IU6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Br5538 protein.
 GN Br5538.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobiaceae;
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005955; BAC50803.1; -;

DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR Pfam; PF05658; Hep_Hag; 12.
 DR Pfam; PF05662; HIM; 3.
 KW Complete proteome.
 SQ SEQUENCE 762 AA; 72776 MW; B2DFEE9A1CB241CE CRC64;
 Query Match 34.2%; Score 67; DB 16; Length 762;
 Best Local Similarity 48.6%; Pred. No. 15;
 Matches 17; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 QY 5 SEADGKGKANARGDKSIAIGDIAQAALGSQSIAG 39
 DB 521 TKVNSAGAAANASGTDALAGNAQAATGSGSIAIG 555
 RESULT 10
 Q9FDA0 PRELIMINARY; PRT; 1265 AA.
 AC Q9FDA0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative outer membrane protein Xada.
 GN XADA.
 OS Xanthomonas oryzae (pv. oryzae).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=64187;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ray S.K., Rajeshwari R., Sonti R.V.;
 RT "A putative outer membrane protein from *Xanthomonas oryzae* pv. *oryzae*
 that is involved in virulence.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF288222; AAG01335.1; -;
 DR InterPro; IPR008640; Hep_Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep_Hag; 25.
 DR Pfam; PF05662; HIM; 4.
 DR Pfam; PF03895; Yada; 1.
 SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;
 Query Match 34.2%; Score 67; DB 2; Length 1265;
 Best Local Similarity 46.2%; Pred. No. 25;
 Matches 12; Conservative 8; Mismatches 6; Indels 0; Gaps 0;
 QY 14 ANARGDKSIAIGDIAQAALGSQSIAG 39
 DB 694 AQATGVSVAIGELSKATGESVAVG 719
 RESULT 11
 Q92KQ7 PRELIMINARY; PRT; 1291 AA.
 ID Q92KQ7;
 AC Q92KQ7;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein R00458.
 GN R00458 OR SMC01708.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,


```

RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591783; CAC41895.1; -.
DR GO: GO:0005509; F:calcium ion binding; IBA.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR008640; Hep Hag.
DR InterPro: IPR008635; HIM.
DR InterPro: IPR005594; Yada.
DR Pfam: PF05658; Hep Hag; 13.
DR Pfam: PF05662; HIM; 6.
DR Pfam: PF03895; Yada; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1291 AA; 127509 MW; 1B0F2A8CD1B1613C CRC64;

Query Match 34.2%; Score 67; DB 16; Length 1291;
Best Local Similarity 44.2%; Pred. No. 26;
Matches 19; Conservative 3; Mismatches 13; Indels 8; Gaps 1;

QY 5 SEADGGKGG-----ANARGKSIAGDIAQALGQSQTAI 39
DB 88 NEADPGASDAITGTDAGRSLSAIGRQAGNEQSIGIG 130

RESULT 12
ID Q8FCB2 PRELIMINARY; PRT; 1778 AA.
AC Q8FCB2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin.
GN C4424.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O6:H1 / CF7073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL: AE016768; AAN82860.1; -.
DR InterPro: IPR008640; Hep Hag.
DR InterPro: IPR008635; HIM.
DR InterPro: IPR005594; Yada.
DR Pfam: PF05658; Hep Hag; 14.
DR Pfam: PF05662; HIM; 15.
DR Pfam: PF03895; Yada; 1.
DR KW Complete proteome.
SQ SEQUENCE 1778 AA; 177678 MW; 98564AA3A797DA20 CRC64;

Query Match 34.2%; Score 67; DB 16; Length 1778;
Best Local Similarity 39.6%; Pred. No. 37;
Matches 19; Conservative 6; Mismatches 11; Indels 12; Gaps 1;

QY 5 SEADGGKGGANRG-----DKSIAIGDIAQALGQSQTAI 40
DB 217 SLAFGRKSTANSTGSLAIGADSSSDNDAIGNKTKQALGWSNALGN 264

RESULT 13
Q9EQT3
ID Q9EQT3 PRELIMINARY; PRT; 261 AA.

```

```

Q9EQT3;
AC 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GTP-binding protein like 1 (Wrch-1).
GN ARHU OR WRCH1 OR MG28K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Daigo Y., Takayama I., Fujino M.A.;
RT "Isolation, mapping, and characterization of a novel murine cDNA,
RT MG28K homologous to the mammalian GTP-binding protein family.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21352806; PubMed=11459829;
RA Tao W., Pennica D., Xu L., Kalejta R.F., Levine A.J.;
RT "Wrch-1, a novel member of the Rho gene family that is regulated by
RT Wnt-1.";
RL Genes Dev. 15:1796-1807(2001).
CC -!- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
DR EMBL: AB051827; BAB18639.1; -.
DR EMBL: AF378088; AAK83341.1; -.
DR HSSP: P21181; 1AM4.
DR MGD; MGI:1316831; Arhu.
DR GO: GO:0003931; F:Rho small monomeric GTPase activity; IMP.
DR GO: GO:0030036; P:actin cytoskeleton organization and biogenesis; IDA.
DR GO: GO:0000082; P:G1/S transition of mitotic cell cycle; IDA.
DR GO: GO:0008360; P:regulation of cell shape; IDA.
DR InterPro: IPR003578; GTPase Rho.
DR InterPro: IPR001806; Ras trnsfrmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00174; RHO; 1.
DR TIGRFAMS; TIGR00231; small GTP; 1.
DR KW GTP-binding; Lipoprotein; Prenylation.
SQ SEQUENCE 261 AA; 28353 MW; 8E4250071E75B14 CRC64;

Query Match 33.7%; Score 66; DB 11; Length 261;
Best Local Similarity 39.5%; Pred. No. 6.1;
Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 1;

QY 1 GTGISADGGKGGANRGKSIAGDIAQALGQSQTAI 38
DB 35 GPGVSGRGRAGGAGRGVCKVLVGD--GAVGKTSLVV 70

RESULT 14
Q9PD63
ID Q9PD63 PRELIMINARY; PRT; 1004 AA.
AC Q9PD63;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Surface-exposed outer membrane protein.
GN XFI516.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9a5c;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colaudo N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:32:39 ; Search time 0.743764 Seconds
(without alignments)
2800.358 Million cell updates/sec

Title: US-09-813-214A-11

Perfect score: 196

Sequence: 1 GIGISEADGGKGNARCKSIAGIDIAQAALGQSQAIGD 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	31.1	377	1 RPOA_CHLMU	Q9p1n4 Chlamydia m
2	61	31.1	455	1 YADA_YEREN	P31489 yersinia en
3	60	30.6	278	1 T2D7_DROME	Q27272 drosophila
4	59.5	30.4	434	1 YADA_YERPS	P10958 yersinia ps
5	58.5	29.8	104	1 CH10_SCHPO	O59804 schizosach
6	57.5	29.3	403	1 CK05_HUMAN	Q9uhf6 homo sapien
7	57.5	29.3	676	1 ICPO_HSVBK	P29836 bovine herp
8	57.5	29.3	710	1 MASZ_PSEFL	O05137 pseudomonas
9	57.5	29.3	725	1 YZ08_MYCTU	O53553 mycobacteri
10	57	29.1	1901	1 NBEA_HUMAN	Q8nfp9 homo sapien
11	57	29.1	2946	1 RPOA_CHLMU	Q46449 chlamydia t
12	56	28.6	377	1 RPOA_CHLMU	P94212 alcaaligenes
13	55.5	28.3	498	1 NDDD_ALCXX	P94212 alcaaligenes
14	55.5	28.3	698	1 EFG_VIBCH	Q9kuz7 vibrio chol
15	54.5	27.8	258	1 BDHA_ALCEU	O9x6u2 alcaaligenes
16	54.5	27.8	730	1 KDGA_MOUSE	O88673 mus musculu
17	54	27.6	245	1 YH77_ARCFU	O28497 archaeoglob
18	54	27.6	370	1 LEU3_BRAJA	Q89x19 bradyrhizob
19	53.5	27.3	331	1 MACS_BOVIN	P12624 bos taurus
20	53.5	27.3	699	1 EFG_HABIN	P43925 haemophilus
21	53.5	27.3	699	1 EFG_VIRPA	Q87145 vibrio para
22	53.5	27.3	699	1 EFG_VIRVU	Q8dgc8 vibrio vuln
23	53.5	27.3	700	1 EFG_PASMU	P57938 pasteurella
24	53.5	27.3	703	1 EFG_ECOLI	P02996 escherichia
25	53.5	27.3	703	1 EFG_SALTY	P26229 salmonella
26	53.5	27.3	734	1 KDGA_PIG	P20192 sus scrofa
27	53	27.0	219	1 PGMB_ECOLI	P77366 escherichia
28	53	27.0	401	1 YK03_CAEEL	P34291 caenorhabdi
29	53	27.0	555	1 CSW_DROSI	Q8isc9 drosophila
30	53	27.0	595	1 PRIM_CHLTR	O84799 chlamydia t
31	52.5	26.8	329	1 HEM2_MYCLE	P46723 mycobacteri
32	52.5	26.8	459	1 COBG_PSEDE	P21637 pseudomonas
33	52.5	26.8	698	1 EFG1_SHEON	Q8ek71 shewanella

ALIGNMENTS

RESULT 1

RPOA_CHLMU
ID_RPOA_CHLMU STANDARD; PRT; 377 AA.
AC Q9PJN4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR TC0794.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Ni9g;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback I., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";
RT Nucleic Acids Res. 28:1397-1406(2000).
RL FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.
CC CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).
CC SUBUNIT: Homodimer. The RNAP catalytic core consists of 2 alpha, 1 beta, 1 beta', and 1 omega subunit (By similarity).
CC DOMAIN: The N-terminal domain is essential for RNAP assembly and basal transcription, whereas the C-terminal domain is involved in interaction with transcriptional regulators and with upstream promoter elements (By similarity).
CC SIMILARITY: Belongs to the RNA polymerase alpha chain family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC EMBL; AR002347; AAF39597.1; ALT_INIT.
CC HSSP; P00574; LCOO.
CC TIGR; TC0794; -.
CC HAMAP; MF 00059; -.
CC InterPro; IPR009025; RBP11-like RNAPo.
CC InterPro; IPR001700; RNA_pol_A_bac_org.
CC Pfam; PF01000; RNA_pol_A_bac_1.
CC Pfam; PF03118; RNA_pol_A_CTD; 1.
CC ProDom; PD001179; RNA_polA_bac_org; 1.
CC SMART; SM00662; RPOLD; 1.

34 52.5 26.8 699 1 EFG_COXBU Q83es7 coxiella bu
35 52.5 26.8 702 1 EFG_BUCAL P57593 buchnera ap
36 52.5 26.8 702 1 EFG_YERPE Q8k948 buchnera ap
37 52.5 26.8 702 1 EFG_YERPE Q8zjb3 yersinia pe
38 52.5 26.8 703 1 EFG_PSEPK Q8efi4 pseudomonas
39 52.5 26.8 705 1 EFG_WIGBR Q8d3h2 wiggleswort
40 52.5 26.8 727 1 KDGA_RAT P51556 rattus norv
41 52 26.5 232 1 GIDB_COREF Q8fsv1 corynebacte
42 52 26.5 299 1 YC68_SULTO Q971w0 sulfolobus
43 52 26.5 348 1 IDI2_LACPL Q88wb6 lactobacill
44 52 26.5 550 1 USHA_SALPU Q9rn37 salmonella
45 52 26.5 550 1 USHA_SALTY P06196 salmonella

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: TAFs are components of the transcription factor IID
 CC (TFIID) complex that are essential for mediating regulation of RNA
 CC polymerase transcription (by similarity).
 CC -!- SUBUNIT: TFIID is composed of TAF binding protein (TBP) and a
 CC number of TBP-associated factors (TAFs).
 CC -!- SIMILARITY: Belongs to the TAF2G family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U06458; AAC47347.1; -.
 CC EMBL: L29540; AAA28488.1; -.
 CC EMBL: AE003506; AAF48767.1; -.
 CC PIR: A49067; A49067.
 CC TRANSFAC: T02125; -.
 CC FlyBase: FBgn000617; e(y)1.
 CC InterPro: IPR007124; Hist.TAF.
 CC InterPro: IPR003162; TFIID-31.
 CC Pfam: PF02291; TFIID-31; 1.
 CC ProDom: PD011023; TFIID-31; 1.
 CC Transcription regulation; Nuclear protein.
 CC DOMAIN 190 197 POLY-THR.
 CC DOMAIN 202 209 POLY-GLY.
 CC DOMAIN 250 256 POLY-GLY.
 CC DOMAIN 267 274 POLY-GLU (ACIDIC).
 CC SEQUENCE 278 AA; 29314 MW; 0EA442C80467001F CRC64;
 KW Transcription regulation; Nuclear protein.
 FT DOMAIN 190 197 POLY-THR.
 FT DOMAIN 202 209 POLY-GLY.
 FT DOMAIN 250 256 POLY-GLY.
 FT DOMAIN 267 274 POLY-GLU (ACIDIC).
 SQ SEQUENCE 278 AA; 29314 MW; 0EA442C80467001F CRC64;
 Query Match 30.6%; Score 60; DB 1; Length 278;
 Best Local Similarity 31.6%; Pred. No. 5.9;
 Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;
 QY 2 IGISEADGGKGNARGDKSIAIGDIAQALGSGSQTIG 39
 DB 198 VGSSEGGGGGGGQEVKSESTGAGGDLKMEVDSDAAV 235
 RESULT 4
 ID YADA YERPS STANDARD; PRT; 434 AA.
 AC P10858;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last annotation update)
 DE Invasin precursor (Outer membrane adhesin).
 GN YADA OR YOPA OR INVA OR YOP1.
 OS *Yersinia pseudotuberculosis*.
 OG Plasmid pIB1.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Yersinia*.
 OX NCBI_TaxID=633;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YPIII;
 RX MEDLINE=88302441; PubMed=3043229;
 RA Rosqvist R., Skurnik M., Wolf-Watz H.;
 RT "Increased virulence of *Yersinia pseudotuberculosis* by two
 RT independent mutations.";
 RL Nature 334:522-525(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YPIII;

RX MEDLINE=89343638; PubMed=2761389;
 RA Skurnik M., Wolf-Watz H.;
 RT "Analysis of the yopA gene encoding the YopI virulence determinants
 of *Yersinia* spp.";
 RL Mol. Microbiol. 3:517-529(1989).
 CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
 CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
 CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
 CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
 CC SURFACE.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X13983; CAA32088.1; -.
 CC PIR: S04534; S04534.
 CC InterPro: IPR008126; Adhesion.
 CC InterPro: IPR008640; Hep Hag.
 CC InterPro: IPR008635; HIM.
 CC InterPro: IPR005594; Yada.
 CC Pfam: PF05658; Hep Hag; 4.
 CC Pfam: PF05662; HIM; 1.
 CC Pfam: PF03895; Yada; 1.
 CC PRINTS: PR01756; OMADHESIN.
 KW Plasmid; Virulence; Signal; Outer membrane.
 FT SIGNAL 1 25 INVASIN
 FT CHAIN 26 434
 FT SEQUENCE 434 AA; 45054 MW; EE2C55FB12B183D4 CRC64;
 Query Match 30.4%; Score 59.5; DB 1; Length 434;
 Best Local Similarity 48.7%; Pred. No. 10;
 Matches 19; Conservative 2; Mismatches 9; Indels 9; Gaps 2;
 QY 10 GKGNARGDK--STAIGDIAQ-----ALGSGSQTIG 39
 DB 94 GAGGLNARAKDPYSIAIGATAEAAPAAVAVGSGSIATG 132
 RESULT 5
 ID CH10 SCHPO STANDARD; PRT; 104 AA.
 AC O59804;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 10 kDa heat shock protein, mitochondrial (HSP10) (10 kDa chaperonin).
 GN HSP10 OR SPCC550.06C.
 OS *Schizosaccharomyces pombe* (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Murgall K., Murphy L., Niblett D., Odell J.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,


```

FT ZN_FING      13      52      RING-TYPE.
FT DOMAIN      284     331      ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE    676 AA; 67701 MW; 9B80683C9BFC65D CRC64;

Query Match      29.3%; Score 57.5; DB 1; Length 676;
Best Local Similarity 44.1%; Pred. No. 27;
Matches 15; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

QY 3 GISEADGGKGGANARGD-KSIAIGDIAQALGSQS 35
DB 120 GGSEAGGGAGGAAGAAGAGAGAGAGAGAGAGGA 153

RESULT 8
EFG_BUCBP
ID EFG_BUCBP STANDARD; PRT; 710 AA.
AC P59451;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Elongation factor G (EF-G).
OS FUSA OR BHP470.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tanames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC EF-G/EF-2 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014017; AAO27176.1; -.
CC HAMAP; MF_00054; -.
CC InterPro; IPR004540; EF-G.
CC InterPro; IPR000795; EF_GTPbind.
CC InterPro; IPR000640; EFG C.
CC InterPro; IPR009022; EFG_III_V.
CC InterPro; IPR005517; EFG_IV.
CC InterPro; IPR004161; EFTU_D2.
CC InterPro; IPR005225; Small_GTP.
CC InterPro; IPR009000; Translat_factor.
CC Pfam; PF00679; EFG C; 1.
CC Pfam; PF03764; EFG IV; 1.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGATNFCT.
CC TIGRFAMs; TIGR00484; EF-G; 1.
CC TIGRFAMs; TIGR00231; small GTP; 1.
CC PROSITE; PS00301; EFATOR GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 710 AA; 79363 MW; 037BA37B190557FD CRC64;

Query Match      29.3%; Score 57.5; DB 1; Length 725;
Best Local Similarity 65.2%; Pred. No. 29;
Matches 15; Conservative 0; Mismatches 5; Indels 3; Gaps 1;

QY 4 ISEADG---GKGGANARGDKSIA 23
DB 148 ISEADGAEKGGYKVRGDKVIA 170

RESULT 10
YZ08_MYCTU
ID YZ08_MYCTU STANDARD; PRT; 1901 AA.
AC O53553;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [4]
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL020222; CAA17745.1; .
DR PIR; F70806; F70806.
DR TubercuList; RV3508;
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30 POTENTIAL
FT CHAIN 31 1901 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT RV3508.
SQ SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match 29.1%; Score 57; DB 1; Length 1901;
Best Local Similarity 59.1%; Pred. No. 85;
Matches 13; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
QY 1 GIGISEADGKGKGNARGDKSI 22
|||: ||||| |||||
Db 1307 GIGGTGGAGGAGGAGDPSI 1328
RESULT 11
ID NBEA_HUMAN STANDARD; PRT; 2946 AA.
AC Q9NF99; Q9HCW8; Q9NSU1; Q9NW98; Q9Y6J1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurobeachin protein (lysosomal trafficking regulator 2) (BCL8B
DE protein).
GN NBEA OR LYST2 OR BCL8B OR KIAA1544.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC TISSUE=Brain, and Spleen;
RX MEDLINE=22150869; PubMed=12160729;
RA Dyomin V.G., Chaganti S.R., Dyomina K., Palanisamy N., Murty V.V.V.S.,
RA Dalla-Favera R., Chaganti R.S.K.;
RT "BCL8 is a novel, evolutionarily conserved human gene family encoding
RT proteins with presumptive protein kinase A anchoring function.";
RL Genomics 80:158-165(2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RA Duisterhoef A., Lauber J., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 606-1118 FROM N.A.
RC TISSUE=Embryonic head;
RA Itogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1919-2946 FROM N.A. (ISOFORM 1).
RX MEDLINE=20450683; PubMed=10957877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
RN [5]
RP SEQUENCE OF 2428-2946 FROM N.A. (ISOFORM 1).
RA Tchernev V.T., McMurtre E.B., Nguyen Q.A., Mishra V.S.,
RA Barbosa M.D.F.S., McIndoe R., Kingmore S.F.;
RT "Identification of LYST2, a brain-specific member of the Chediak-
RT Higaishi syndrome gene family.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 2150-2563.
RX MEDLINE=22220051; PubMed=12234919;
RA Jogl G., Shen Y., Gebauer D., Li J., Wiegmann K., Kashkar H.,
RA Krenke M., Tong L.;
RT "Crystal structure of the BEACH domain reveals an unusual fold and
RT extensive association with a novel PH domain.";
RL EMBO J. 21:4785-4795(2002).
CC -1- FUNCTION: Binds to type II regulatory subunits of protein kinase A
CC and anchors/targets them to the membrane. May anchor the kinase to
CC cytoskeletal and/or organelle-associated proteins (By
CC similarity).
CC -1- SUBUNIT: Interacts with RII subunit of PKA (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8NFP9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NFP9-2; Sequence=VSP_050538, VSP_050539;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Predominant in many brain structures. Also
CC expressed at medium levels in spleen, thymus, prostate, testis and
CC ovary. Low level expression is seen in heart, kidney, pancreas,
CC skeletal muscle and intestine.
CC -1- DOMAIN: RII-alpha binding site, predicted to form an amphipathic
CC helix, could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer (By
CC similarity).
CC -1- SIMILARITY: Belongs to the neurobeachin family.
CC -1- SIMILARITY: Contains 1 BEACH domain.
CC -1- SIMILARITY: Contains 5 WD repeats.
CC -1- CAUTION: Ref.3 sequence differs from that shown due to a stop
CC codon in position 762.
CC -1- CAUTION: Ref.5 sequence differs from that shown due to a
CC frameshift in position 2900.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

AC P94212;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE N-acyl-D-aspartate decarboxylase (EC 3.5.1.83) (N-acyl-D-aspartate
 DE amidohydrolase).
 OS Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxydans).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Achromobacter.
 CX NCBI_TaxID=515;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=A-6;
 RC Wakayama M., Watanabe E., Takenaka Y., Miyamoto Y., Tau Y., Sakai K.,
 RA Moriguchi M.;
 RA "Cloning, expression, and nucleotide sequence of the N-acyl-D-
 RT aspartate amidohydrolase gene from Alcaligenes xylosoxydans subsp.
 RT xylosoxydans A-6.";
 RL J. Ferment. Bioeng. 80:311-317(1995).
 RN [2]
 RN CHARACTERIZATION.
 RP STRAIN=A-6;
 RC MEDLINE=93372486; PubMed=7763985;
 RX Moriguchi M., Sakai K., Katsuno Y., Maki T., Wakayama M.;
 RA "Purification and characterization of novel N-acyl-D-aspartate
 RT amidohydrolase from Alcaligenes xylosoxydans subsp. xylosoxydans
 RT A-6.";
 RL Biosci. Biotechnol. Biochem. 57:1145-1148(1993).
 CC -!- CATALYTIC ACTIVITY: N-acyl-D-aspartate + H(2)O = carboxylate + D-
 CC aspartate.
 CC -!- COFACTOR: Zinc.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the N-acyl-D-amino-acid deacylase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D45919; BAA08350.1; -;
 DR InterPro; IPR006680; Amidohydro.1.
 DR InterPro; IPR005847; Pept M38 regn.
 DR Pfam; PF01979; Amidohydro.1; 1.
 DR ProDom; PD000518; Urease; 1.
 DR Hydrolyase; Zinc.
 SQ SEQUENCE 498 AA; 53584 MW; EAF8662891F9F1B CRC64;

Query Match 28.3%; Score 55.5; DB 1; Length 498;
 Best Local Similarity 40.0%; Pred. No. 34;
 Matches 14; Conservative 4; Mismatches 12; Indels 5; Gaps 1;
 QY 3 GISEADGGGGAN-----ARGDKSIAIGDIAQALG 32
 DB 20 GATLDGGGPGARQGLAVRGGRIVALGDFAHAPG 54

RESULT 14

EFQ_VIBCH
 ID EFQ_VIBCH STANDARD; PRT; 698 AA.
 AC Q9KUZ7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUSA OR VC0361.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 CX NCBI_TaxID=666;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=E1 Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
 CC the nascent protein chain from the A-site to the P-site of the
 CC ribosome.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
 CC EF-G/EF-2 subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AE004124; AAP93534.1; -;
 DR PIR; C82332; C82332.
 DR HSSP; P13551; LELO.
 DR TIGR; VC0361; -;
 DR HAMAP; MF 00054; -; 1.
 DR InterPro; IPR004540; EF-G.
 DR InterPro; IPR000795; EF GTPbind.
 DR InterPro; IPR000640; EFG_C.
 DR InterPro; IPR009022; EFG_III_V.
 DR InterPro; IPR005517; EFG_IV.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR005225; Small_GTP.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF00679; EFG_C; 1.
 DR Pfam; PF03764; EFG_IV; 1.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATNFCT.
 DR TIGRFAMS; TIGR00484; EF-G; 1.
 DR TIGRFAMS; TIGR00231; small_GTP; 1.
 DR PROSITE; PS00301; EFACOR GTP; 1.
 KW Elongation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT NP_BIND 17 24 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SQ SEQUENCE 698 AA; 76927 MW; F44FDA4DB8FC4ECC CRC64;

Query Match 28.3%; Score 55.5; DB 1; Length 698;
 Best Local Similarity 46.4%; Pred. No. 48;
 Matches 13; Conservative 4; Mismatches 10; Indels 1; Gaps 1;
 QY 14 ANARGD-KSIAIGDIAQALGSGSIAIGD 40
 DB 366 ANKRDEIKRAGDIAAIGLKDVITGD 393

RESULT 15
 BDHA_ALCEU
 ID BDHA_ALCEU STANDARD; PRT; 258 AA.
 AC Q9X6U2;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE D-beta-hydroxybutyrate dehydrogenase (EC 1.1.1.30) (BDH)

```

DE (3-hydroxybutyrate dehydrogenase) (3-HBDH).
GN HBDH1
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H16 / DSM 428 / ATCC 17699;
RA Kim J.W., Kang D.G., Rha E.G.;
RT "Cloning and sequencing of the gene for beta-hydroxybutyrate
RL dehydrogenase from Ralstonia eutropha.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: (R)-3-hydroxybutanoate + NAD(+) = acetoacetate
CC + NADH.
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
CC (SDR) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF145230; AAD33952.1; -.
DR HSBP; O70351; IE8W.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PRO0080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 8 32 NAD (BY SIMILARITY).
FT ACT_SITE 153 153 BY SIMILARITY.
SQ SEQUENCE 258 AA; 27014 MW; 269A06D6CD97FAEF CRC64;
Query Match 27.8%; Score 54.5; DB 1; Length 258;
Best Local Similarity 38.1%; Pred. No. 23;
Matches 16; Conservative 6; Mismatches 15; Indels 5; Gaps 2;
QY 1 GIGISEADGKGG--ANARGDKSIAIGDIAQALGQSIAIG 39
Db 17 GLGIKALAAQGANIIVNGFGDADAQAEIAQA--GGIRVG 56

```

Search completed: September 21, 2004, 23:43:47
 Job time : 1.74376 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:33:49 ; Search time 1.32426 Seconds
(without alignments)
2905.511 Million cell updates/sec

```

Title: US-09-813-214A-11
Perfect score: 196
Sequence: 1 GIGISEADGGKGGANRGPKSTAGDTAOLGSOSTAIGD 40

```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	67.5	34.4	1588	2	A86036	probable adhesin Z
2	67.5	34.4	1588	2	H9188	probable adhesin E
3	66	33.7	1004	2	C82672	surface-exposed ou
4	64.5	32.9	1190	2	A2615	surface protein XF
5	63	32.1	615	2	H70589	hypothetical glyci
6	63	32.1	694	2	F70868	hypothetical glyci
7	61.5	31.4	2342	2	T18200	fatty-acid synthas
8	61	31.1	394	2	B81663	DNA-directed RNA p
9	61	31.1	455	2	S04912	yopA protein - Yer
10	61	31.1	658	2	A00110	probable surface p
11	60.5	30.9	2059	2	D83671	surface protein XF
12	60	30.6	278	2	A49067	transcription init
13	60	30.6	327	2	T00797	hypothetical prote
14	60	30.6	621	2	T45851	conserved hypothet
15	59.5	30.4	434	2	S04534	invasin precursor
16	58.5	29.8	104	2	T41381	Chaperonins 10 Kd
17	58	29.6	1176	2	T18042	ice nucleation pro
18	57.5	29.3	436	2	T36478	probable solute-bi
19	57.5	29.3	676	1	EDB822	immediate-early pr
20	57.5	29.3	1538	2	H70846	hypothetical glyci
21	57.5	29.3	1660	2	A70869	hypothetical glyci
22	57	29.1	138	2	C87389	hypothetical prote
23	57	29.1	422	2	S04911	yopA protein - Yer
24	57	29.1	505	2	AC3486	cell surface prote
25	57	29.1	1398	2	C87448	hypothetical prote
26	57	29.1	1737	2	A59235	unconventional myo
27	57	29.1	1901	2	F70806	hypothetical glyci
28	56.5	28.8	882	2	D70812	hypothetical glyci
29	56	28.6	365	2	AB3486	cell surface prote

ALIGNMENTS

RESULT 1

A86036
Probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: A86036
C/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dinalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: A86036
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1588 <STO>
A/Cross-references: GB:AF005174; NID:gl2518349; PIDN:AA658749.1; GSPDB:GN00145; UWGP:Z5029
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: Z5029

```

Query Match          34.4%; Score 67.5; DB 2; Length 1588;
Best Local Similarity 33.3%; Pred.No.6.4;
Matches 18; Conservative      8; Mismatches    9; Indels   19; Gaps    2;

Qy      5 SEADG-----GKG-----GANARGDKGIAGIDIAQAALGSQSIAIG 39
       :|||         :|||         :|||         :|||         :|||
Db      281 _NKADGVDAIALGNQSQRGLNTIALGTASNATGDKSLALGNSSSANGINSVALG 334

RESULT 2
H91188
probable adhesin ECs4480 [similarity] - Escherichia coli (strain O157:H7, subst
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
C/Accession: H91188
R/Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; I
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

```

DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: H91188
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1588 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BAB37903.1; PID:gl3363955; GSPDB:GNC00154
 A;Experimental source: strain O157:H7, substrain RIMD 050952
 C;Genetics:
 A;Gene: ECs4480

Query Match	34.4%	Score 67.5;	DB 2;	Length 1588;
Best Local Similarity	33.3%	Pred. No. 6.4;		
Matches 18;	Conservative	8;	Mismatches 9;	Indels 19;
				Gaps 2;

Query Match 32.1%; Score 63; DB 2; Length 694;
 Best Local Similarity 44.4%; Pred. No. 9.7;
 Matches 20; Conservative 3; Mismatches 14; Indels 8; Gaps 3;
 QY 1 GIGISEADGGKGGANARGDKSTAIAGDIAQAL-----GSQSIAIG 39
 Db 296 GGGGTAAAGG-GGGNG-GDGGVQAQGDIAAFGGDGGSDGVVAG 338
 RESULT 7
 T18200
 fatty-acid synthase homolog p260 - silkworm
 C:Species: Bombyx mori (silkworm)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
 C:Accession: T18200
 R:Ueno, K.
 submitted to the EMBL Data Library, August 1996
 A:Description: P260/270 expressed in embryonic abdominal leg. cells of Bombyx mori can tr
 A:Reference number: Z18815
 A:Accession: T18200
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2342 <UN>
 A:Cross-references: EMBL:U67866; NID:G2058457; PID:G2058458; PIDN:AA853257.1
 C:Superfamily: rat fatty-acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hom
 yrolyase homology; short-chain alcohol dehydrogenase homology; [acyl-carrier-protein] S-
 C:Keywords: carrier protein
 Query Match 31.4%; Score 61.5; DB 2; Length 2342;
 Best Local Similarity 40.0%; Pred. No. 48;
 Matches 18; Conservative 5; Mismatches 13; Indels 9; Gaps 2;
 QY 1 GIGISEADG-----GKGANARGDKSTAIAGDIAQALGSQSIAIG 39
 Db 322 GAATAEADGKLEAIGKVFAN---DKTIKVGCVKSNMGSSAAAG 363
 RESULT 8
 B81663
 DNA-directed RNA polymerase, alpha chain TC0794 [imported] - Chlamydia muridarum (strain
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
 C:Accession: B81663
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: B81663
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-394 <TET>
 A:Cross-references: GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF39597.1; PID:g719082
 A:Experimental source: strain Nigg (MoPn)
 C:Genetics:
 A:Gene: TC0794
 C:Superfamily: DNA-directed RNA polymerase alpha chain
 Query Match 31.1%; Score 61; DB 2; Length 394;
 Best Local Similarity 32.4%; Pred. No. 9.5;
 Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 QY 4 TSEADGGKGGANARGDKSTAIAGDIAQALGSQSIAIGD 40
 Db 123 LQDCEGGRASQKLRTATISDASDLAAAGGQKAITLGD 159
 RESULT 9
 S04912
 yopA protein - Yersinia enterocolitica plasmid pYV6471/76
 C:Species: Yersinia enterocolitica
 C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S04912
 R:Skurnik, M.; Wolf-Watz, H.
 Mol. Microbiol. 3, 517-529, 1989
 A:Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
 A:Reference number: S04910; MUID:89343638; PMID:2761389
 A:Accession: S04912
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-455 <SKU>
 A:Cross-references: EMBL:X13882; NID:G48606; PIDN:CAA32086.1; PID:G48607
 C:Genetics:
 A:Genome: plasmid pYV6471/76
 Query Match 31.1%; Score 61; DB 2; Length 455;
 Best Local Similarity 46.9%; Pred. No. 11;
 Matches 15; Conservative 6; Mismatches 9; Indels 2; Gaps 1;
 QY 10 GKGG--ANARGDKSTAIAGDIAQALGSQSIAIG 39
 Db 59 GAGGLNASAKGIHSTAIAGTAAEAAKGAAVAG 90
 RESULT 10
 AH0110
 probable surface protein (partial) YPO0902 [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C:Accession: AH0110
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 I., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AH0001; MUID:21470413; PMID:11586360
 A:Accession: AH0110
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-658 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:GL5978974; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO0902
 Query Match 31.1%; Score 61; DB 2; Length 658;
 Best Local Similarity 36.7%; Pred. No. 16;
 Matches 18; Conservative 6; Mismatches 15; Indels 10; Gaps 2;
 QY 1 GIGI-----SEADGGKGG--GANARGDKSTAIAGDIAQALGSQSIAIG 39
 Db 167 GVGIGNTALVGAATGGTGAIGTGQTAAAGATAIGSAQAQAQASLALG 215
 RESULT 11
 D82671
 surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82671
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82671
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2059 <SIM>
 A:Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN001.
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.U. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XFI529

Query Match 30.9%; Score 60.5; DB 2; Length 2059;
Best Local Similarity 42.5%; Pred. No. 55;
Matches 17; Conservative 3; Mismatches 15; Indels 5; Gaps 1;

QY 5 SEADGGKGGAN-----ARGDKSIAIGDIAQAALGSQSIAG 39
Db 1722 STNDGGTGGNDGDGATGSKAIAAGVGTQAGSGEAAVG 1761

RESULT 12

A49067
transcription initiation factor IID chain p42 - fruit fly (*Drosophila melanogaster*)
N;Alternate names: TATA-binding protein (TBP)-associated factor TAFII40; TFIID complex C;Species: *Drosophila melanogaster*
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Nov-2000
C;Accession: A49067; S42221
R;Goodrich, J.A.; Hoey, T.; Thut, C.J.; Admon, A.; Tjian, R.
Cell 75, 519-530, 1993
A;Title: *Drosophila* TAF-II40 interacts with both a VP16 activation domain and the basal A;Reference number: A49067; MUID:94037099; PMID:8221891
A;Accession: A49067
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-278 <GO>
A;Cross-references: GB:I29540; NID:9463048; PID:9463049
R;Okubo, T.; Gong, D.W.; Wootton, J.C.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.
Nature 367, 484-487, 1994
A;Title: Molecular cloning of *Drosophila* TFIID subunits.
A;Reference number: S42220; MUID:94150630; PMID:7545910
A;Accession: S42221
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-278 <XOK>
A;Cross-references: EMBL:U06458; NID:9458679; PIDN:AAC47347.1; PID:9458680
C;Genetics:
A;Gene: FlyBase:Taf40
C;Keywords: transcription initiation

Query Match 30.6%; Score 60; DB 2; Length 278;
Best Local Similarity 31.8%; Pred. No. 8.9;
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 2 IGISEADGGKGGANARGDKSIAIGDIAQAALGSQSIAG 39
Db 198 VGSSGGSGGGGGQGVKSESGAGDLKMEVDSDAAAVG 235

RESULT 13

T00797
hypothetical protein At2g32710 [imported] - *Arabidopsis thaliana*
N;Alternate names: hypothetical protein F24L7.15
C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C;Accession: T00797; E84736
R;Rounsley, S.D.; Lin, X.D.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul submitted to the EMBL Data Library, February 1998
A;Description: *Arabidopsis thaliana* chromosome II BAC F24L7 genomic sequence.
A;Reference number: Z14204
A;Accession: T00797

A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-327 <ROU>
A;Cross-references: EMBL:AC003974; NID:92914688; PID:92914702
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84736
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-327 <STO>
A;Cross-references: GB:AE002093; NID:92914702; PIDN:AAC04492.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32710; F24L7.15
A;Map position: 2
A;Introns: 193/2
C;Superfamily: *Arabidopsis thaliana* hypothetical protein F24L7.15

Query Match 30.8%; Score 60; DB 2; Length 327;
Best Local Similarity 33.3%; Pred. No. 10;
Matches 12; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY 3 GISEADGGKGGANARGDKSIAIGDIAQAALGSQSIAG 38
Db 12 GAGAGAGGGGGGGGESSIALMDVWVSPSSSLGV 47

RESULT 14

T46851
conserved hypothetical protein ybaU [imported] - *Rhodobacter sphaeroides*
C;Species: *Rhodobacter sphaeroides*
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C;Accession: T46851
R;Mackenzie, C.; Simmons, A.E.; Kaplan, S.
Genetics 153, 525-538, 1999
A;Title: Multiple chromosomes in bacteria. The Yin and Yang of trp gene localization in A;Reference number: Z24108; MUID:99442363; PMID:10511537
A;Accession: T46851
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-621 <MAC>
A;Cross-references: EMBL:AF108766; NID:94185542; PIDN:AAD09115.1; PID:94185543
A;Experimental source: strain 2.4.1
C;Genetics:
A;Gene: ybaU

Query Match 30.6%; Score 60; DB 2; Length 621;
Best Local Similarity 33.3%; Pred. No. 20;
Matches 13; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 1 GIGISBADGGKGGANARGDKSIAIGDIAQAALGSQSIAG 39
Db 34 GFGVTFGGGLSTVGSVGDRIITGAVARALQEMRALG 72

RESULT 15

S04534
invasin precursor - *Yersinia pseudotuberculosis* plasmid pIBI
C;Species: *Yersinia pseudotuberculosis*
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S04534; S04910
R;Rosqvist, R.; Skurnik, M.; Wolf-Watz, H.
Nature 334, 522-525, 1998
A;Title: Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations A;Reference number: S04534; MUID:88302441; PMID:3043229
A;Accession: S04534
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <ROS>

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:24:53 ; Search time 4.11791 Seconds
(without alignments)
2744.569 Million cell updates/sec

Title: US-09-813-214A-11

Perfect score: 196

Sequence: 1 GIGISEADGGKGGANARGDKSIAIGDIAQAALGQSQIAIGD 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	196	100.0	40	4	AAE00705	AAE00705	N-termina
2	196	100.0	2122	6	ABU08784	ABU08784	Moraxella
3	196	100.0	2123	4	AAE00701	AAE00701	Moraxella
4	190	96.9	43	2	AAW32274	AAW32274	M. catarr
5	190	96.9	43	4	AAE00703	AAE00703	N-termina
6	121	61.7	24	2	AAW32276	AAW32276	M. catarr
7	121	61.7	24	4	AAE00702	AAE00702	Internal
8	86.5	44.1	1946	6	ABU35023	ABU35023	Protein e
9	84	42.9	1992	2	AAW04505	AAW04505	Moraxella
10	84	42.9	1992	4	AAE69137	AAE69137	M. catarr
11	84	42.9	1992	4	AAE69133	AAE69133	M. catarr
12	84	42.9	2047	4	AAE69134	AAE69134	M. catarr
13	84	42.9	2053	4	AAE69135	AAE69135	M. catarr
14	84	42.9	2139	6	ABP71294	ABP71294	M. catarr
15	84	42.9	2314	4	AAE69136	AAE69136	M. catarr
16	75	38.3	2265	6	ABU17199	ABU17199	Protein e
17	75	38.3	2504	6	ADA34534	ADA34534	Acinetoba
18	74	37.8	3073	6	ABU21223	ABU21223	Protein e
19	67.5	34.4	1588	7	ADC01413	ADC01413	Enterohae
20	67	34.2	1778	4	ABE52677	ABE52677	Escherich
21	67	34.2	2042	2	AAW56319	AAW56319	Haemophil
22	66	33.7	261	2	AAI13486	AAI13486	Mouse clo
23	66	33.7	435	6	ABP76894	ABP76894	N. gonorr
24	66	33.7	2712	6	ABU39146	ABU39146	Protein e
25	65.5	33.4	270	6	ABU19868	ABU19868	Protein e

ALIGNMENTS

RESULT 1

AAE00705

ID AAE00705 standard; peptide; 40 AA.

XX AAE00705;

DT 02-JUL-2001 (first entry)

XX

DE N-terminal #2 of M. catarrhalis outer membrane protein-106 (OMP106).

XX

KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;

KW

KW bacterial infection; immunogen; cytotoxic; antibiotic;

KW

XX passive immunisation.

XX

OS Moraxella catarrhalis.

XX

FN US6214981-B1.

XX

PD 10-APR-2001.

XX

PF 12-NOV-1997; 97US-00968685.

XX

PR 03-MAY-1996; 96US-00642712.

XX

(ANTE-) ANTEX BIOLOGICS INC.

PI

Tucker K, Plosila L, Tillman UP;

XX

WPI; 2001-281002/29.

XX

Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane

PT

protein-106 polypeptide, useful for diagnosis of bacterial infections and

PT

as vaccine against Moraxella catarrhalis infection of mammals.

XX

Example; Col 31; 49pp; English.

XX

The present sequence is N-terminal of haemagglutinating Moraxella
catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a
therapeutic and prophylactic vaccine against M. catarrhalis infections of
mammals. It is used for diagnosis of bacterial infections and as reagents
for clinical or medical diagnosis of M. catarrhalis infections and for
scientific research on the properties of M. catarrhalis, virulence and
infectivity of M. catarrhalis. It is also used as a probe to identify the
presence of M. catarrhalis in biological specimens and to identify other
bacteria that encode a polypeptide related to M. catarrhalis OMP106.
OMP106-derived polypeptides are used as ligands to detect antibodies
elicited in response to M. catarrhalis infections and also as immunogens
for inducing M. catarrhalis-specific antibodies which are useful in

26	65	33.2	1461	6	ABU47415	Protein e
27	64	32.7	1099	6	ABU20296	Protein e
28	64	32.7	1400	6	ABU45394	Protein e
29	63	32.1	505	6	ABU34417	Protein e
30	63	32.1	615	6	ABU36862	Protein e
31	63	32.1	694	6	ABU36802	Protein e
32	63	32.1	1129	6	ABU20124	Protein e
33	61	31.1	394	6	ABU27363	Protein e
34	61	31.1	2039	2	AAW56322	Haemophil
35	60	30.6	278	2	AAW56490	TATA-bind
36	60	30.6	278	2	AAW06081	Drosophil
37	60	30.6	278	2	AAW25025	TATA-bind
38	60	30.6	278	4	ABB62342	Arabidops
39	60	30.6	289	3	AAE27262	Arabidops
40	60	30.6	289	3	AAE40485	Arabidops
41	60	30.6	289	5	AAE25109	Arabidops
42	60	30.6	289	5	AAU72581	Arabidops
43	60	30.6	327	3	AAE27261	Arabidops
44	60	30.6	449	4	ABB66828	Drosophil
45	59	30.1	462	6	ABR57117	ScnecopFCA

CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis
 XX
 SQ Sequence 40 AA;

Query Match 100.0%; Score 196; DB 4; Length 40;
 Best Local Similarity 100.0%; Pred. No. 6.7e-18;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 GIGISEADGKGANARGDKSIAIGDIAQALGQSSTAIGD 40
 |||||
 Db 1 GIGISEADGKGANARGDKSIAIGDIAQALGQSSTAIGD 40

RESULT 2
 ABU08784
 ID ABU08784 standard; protein; 2122 AA.
 XX
 AC ABU08784;
 XX
 DT 28-MAY-2003 (first entry)
 XX
 DE Moraxella catarrhalis outer membrane protein, OMP106.
 XX
 KW Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;
 KW vaccine.
 XX
 OS Moraxella catarrhalis.

XX Key Location/Qualifiers
 FH Peptide 1..68
 FT /label= signal_sequence
 FT 69..2122
 FT /label= Mature_OMP106
 FT 69..111
 FT /label= N-terminal sequence of mature OMP106
 FT /note= "Specifically claimed in claim 9"
 FT Misc-difference 779..880
 FT /note= "Encoded by ATCTAAGGC"

XX US2002177200-A1.
 PN
 XX
 PD 28-NOV-2002.
 XX
 XX 20-MAR-2001; 2001US-00813214.
 PF
 XX
 PR 03-MAY-1996; 96US-00642712.
 PR 12-NOV-1997; 97US-00968685.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L;
 XX
 XX WPI; 2003-328486/31.
 DR
 DR N-PSDB; ABX93525.
 XX
 XX Novel outer membrane polypeptide of Moraxella catarrhalis, useful for
 PT producing an immune response in an animal, and as ligands to detect
 PT antibodies elicited in response to Moraxella infections.
 XX
 PS Claim 33; Page 28-34; 5lpp; English.
 XX
 XX The invention relates to an isolated or substantially pure outer membrane
 CC polypeptide of Moraxella catarrhalis (OMP106 polypeptide). The
 CC polypeptide and its peptide fragments are useful for producing an immune
 CC response in an animal, and as ligands to detect antibodies elicited in
 CC response to M. catarrhalis infections. The polypeptide and its peptide
 CC fragment are further useful as active ingredients in vaccines against M.
 CC catarrhalis infections. The polypeptide is useful to prepare antibodies.
 CC An antibody binding the polypeptide is useful in immunoassays to detect
 CC M. catarrhalis in biological specimens, and also in passive immunisations
 CC against M. catarrhalis infections. An antibody binding the polypeptide is
 CC also useful to facilitate isolation and purification of the polypeptide

CC and its peptide fragment, and as probes for identifying clones in
 CC expression libraries that have inserts encoding the polypeptide and
 CC peptide fragment. An antibody binding the polypeptide is also useful to
 CC diagnose M. catarrhalis infections. The polynucleotide encoding the
 CC polypeptide is useful as a probe to identify the presence of M.
 CC catarrhalis in biological specimens by hybridisation or PCR amplification
 CC and also to detect other bacteria that might encode a polypeptide related
 CC to M. catarrhalis OMP106. The polypeptide, an antibody binding the
 CC polypeptide, its peptide fragment and the polynucleotide encoding the
 CC polypeptide are useful as reagents for clinical medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of M.
 CC pathogenicity, virulence and infectivity of M. catarrhalis, as well as
 CC host defence mechanisms. The present sequence represents the amino acid
 CC sequence of M. catarrhalis outer membrane protein (OMP)-106

XX SQ Sequence 2122 AA;

Query Match 100.0%; Score 196; DB 6; Length 2122;
 Best Local Similarity 100.0%; Pred. No. 5.7e-16;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GIGISEADGKGANARGDKSIAIGDIAQALGQSSTAIGD 40
 |||||
 Db 68 GIGISEADGKGANARGDKSIAIGDIAQALGQSSTAIGD 107

RESULT 3
 AAE00701
 ID AAE00701 standard; protein; 2123 AA.
 XX
 AC AAE00701;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE Moraxella catarrhalis outer membrane protein-106 (OMP106).
 XX
 KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.

XX US6214981-B1.
 PN
 XX
 PD 10-APR-2001.
 XX
 PF 12-NOV-1997; 97US-00968685.
 PR 03-MAY-1996; 96US-00642712.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L, Tillman UF;
 XX
 XX WPI; 2001-281002/29.
 DR
 DR N-PSDB; AAD04029.

XX Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.

XX Claim 7; Col 53-64; 49pp; English.

XX The present sequence is haemagglutinating Moraxella catarrhalis outer
 CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic and
 CC prophylactic vaccine against M. catarrhalis infections of mammals. It is
 CC used for diagnosis of bacterial infections and as reagents for clinical
 CC or medical diagnosis of M. catarrhalis infections and for scientific
 CC research on the properties of pathogenicity, virulence and infectivity of
 CC M. catarrhalis. It is also used as a probe to identify the presence of M.
 CC catarrhalis in biological specimens and to identify other bacteria that
 CC encode a polypeptide related to M. catarrhalis OMP106. OMP106-derived
 CC polypeptides are used as ligands to detect antibodies elicited in

CC response to M. catarrhalis infections and also as immunogens for inducing
 CC M. catarrhalis-specific antibodies which are useful in immunoassays to
 CC detect M. catarrhalis in biological specimens. Cytotoxic antibodies are
 CC useful in passive immunisations against M. catarrhalis

XX SQ Sequence 2123 AA;
 Query Match 100.0%; Score 196; DB 4; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 5.7e-16; Mismatches 0; Indels 0; Gaps 0;
 Matches 40; Conservative 0;

QY 1 GIGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGD 40
 |||||
 DB 68 GIGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGD 107

RESULT 4
 AAW32274
 ID AAW32274 standard; peptide; 43 AA.
 XX AC AAW32274;
 XX DT 08-MAY-1998 (first entry)
 XX DE M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.
 XX KW Outer membrane protein-106; OMP106; vaccine; immune response;
 XX KW cytotoxic antibody; Moraxella catarrhalis.
 XX OS Moraxella catarrhalis.
 XX PN WO9741731-A1.
 XX PD 13-NOV-1997.
 XX PF 28-APR-1997; 97WO-US007679.
 XX PR 03-MAY-1996; 96US-00642712.
 XX PA (ANTE-) ANTEX BIOLOGICS INC.
 XX PI Tucker K, Plosila L;
 XX DR WPI; 1997-558601/51.
 XX PT Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 XX PT vaccines for producing immune responses against M. catarrhalis.
 XX PS Claim 9; Page 23; 78pp; English.

CC This is a peptide fragment of a novel outer membrane protein-106 (OMP106)
 CC polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella
 CC catarrhalis, an haemagglutinating cultivar. The peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of M.
 CC catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis

XX SQ Sequence 43 AA;
 Query Match 96.9%; Score 190; DB 2; Length 43;
 Best Local Similarity 100.0%; Pred. No. 4.2e-17; Mismatches 0; Indels 0; Gaps 0;
 Matches 39; Conservative 0;

QY 2 IGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGD 40
 |||||
 DB 1 IGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGD 39

RESULT 5
 AAE00703
 ID AAE00703 standard; peptide; 43 AA.

XX AAE00703;
 AC 02-JUL-2001 (first entry)
 DT N-terminal #1 of M. catarrhalis outer membrane protein-106 (OMP106).
 XX DE Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX OS Moraxella catarrhalis.
 XX PN US6214981-B1.
 XX PD 10-APR-2001.
 XX PF 12-NOV-1997; 97US-00968685.
 XX PR 03-MAY-1996; 96US-00642712.
 XX PA (ANTE-) ANTEX BIOLOGICS INC.
 XX PI Tucker K, Plosila L, Tillman UF;
 XX DR WPI; 2001-281002/29.

XX Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 XX protein-106 polypeptide, useful for diagnosis of bacterial infections and
 XX as vaccine against Moraxella catarrhalis infection of mammals.
 XX PS Claim 2; Col 31; 49pp; English.

CC The present sequence is N-terminal of haemagglutinating Moraxella
 CC catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a
 CC therapeutic and prophylactic vaccine against M. catarrhalis infections of
 CC mammals. It is used for diagnosis of bacterial infections and as reagents
 CC for clinical or medical diagnosis of M. catarrhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarrhalis. It is also used as a probe to identify the
 CC presence of M. catarrhalis in biological specimens and to identify other
 CC bacteria that encode a polypeptide related to M. catarrhalis OMP106.
 CC OMP106-derived polypeptides are used as ligands to detect antibodies
 CC elicited in response to M. catarrhalis infections and also as immunogens
 CC for inducing M. catarrhalis-specific antibodies which are useful in
 CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis

XX SQ Sequence 43 AA;
 Query Match 96.9%; Score 190; DB 4; Length 43;
 Best Local Similarity 100.0%; Pred. No. 4.2e-17; Mismatches 0; Indels 0; Gaps 0;
 Matches 39; Conservative 0;

QY 2 IGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGD 40
 |||||
 DB 1 IGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGD 39

RESULT 6
 AAW32276
 ID AAW32276 standard; peptide; 24 AA.
 XX AC AAW32276;
 XX DT 08-MAY-1998 (first entry)

XX DE M. catarrhalis outer membrane protein (OMP)-106 partial peptide fragment.
 XX KW Outer membrane protein-106; OMP106; vaccine; immune response;
 XX KW cytotoxic antibody; Moraxella catarrhalis; primer; probe.
 XX OS Moraxella catarrhalis.

XX WO9741731-A1.
 XX
 PD 13-NOV-1997.
 XX
 XX 28-APR-1997; 97WO-US007679.
 XX
 XX 03-MAY-1996; 96US-00642712.
 XX
 XX (ANTE-) ANTEX BIOLOGICS INC.
 XX
 XX Tucker K, Plosila L;
 XX
 XX WPI; 1997-558601/51.
 XX
 XX N-PSDB; AAT86522.
 XX
 XX Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 PT vaccines for producing immune responses against M. catarrhalis.
 XX
 XX Disclosure; Page 58; 78pp; English.
 XX
 XX This is a partial sequence of an amino terminal peptide fragment of a
 CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an
 CC outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating
 CC cultivar. The encoding DNA can be used as a 5' primer for PCR
 CC amplification of a full length OMP106 DNA. The DNA fragment can also be
 CC used as a probe for screening M. catarrhalis genomic libraries for OMP106
 CC polypeptide coding sequences. The encoded peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of M.
 CC catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis
 XX
 XX Sequence 24 AA;
 QY 6 EADGGKGGANARGDKSIAIGDIAQ 29
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24
 XX
 Query Match 61.7%; Score 121; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 6 EADGGKGGANARGDKSIAIGDIAQ 29
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24
 XX
 RESULT 7
 AAE00702
 ID AAE00702 standard; peptide; 24 AA.
 XX
 XX AAE00702;
 XX
 XX 02-JUL-2001 (first entry)
 XX
 XX Internal fragment of N-terminal outer membrane protein-106 (OMP106).
 XX
 XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 XX Moraxella catarrhalis.
 XX
 XX US6214981-B1.
 XX
 XX 10-APR-2001.
 XX
 XX 12-NOV-1997; 97US-00968685.
 XX
 XX 03-MAY-1996; 96US-00642712.
 XX
 XX (ANTE-) ANTEX BIOLOGICS INC.
 XX
 XX Tucker K, Plosila L, Tillman UP;
 XX

DR WPI; 2001-281002/29.
 DR N-PSDB; AAD04030.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 XX Example; Col 43-44; 49pp; English.
 XX
 XX The present sequence is the internal fragment of N-terminal
 CC haemagglutinating Moraxella catarrhalis outer membrane protein-106
 CC (OMP106). This sequence is used to design a probe and a 5' PCR primer.
 CC The OMP106 is used as a therapeutic and prophylactic vaccine against M.
 CC catarrhalis infections of mammals. It is used for diagnosis of bacterial
 CC infections and as reagents for clinical or medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also
 CC used as a probe to identify the presence of M. catarrhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used as
 CC ligands to detect antibodies elicited in response to M. catarrhalis
 CC infections and also as immunogens for inducing M. catarrhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarrhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarrhalis
 XX
 XX Sequence 24 AA;
 QY 6 EADGGKGGANARGDKSIAIGDIAQ 29
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24
 XX
 Query Match 61.7%; Score 121; DB 4; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 6 EADGGKGGANARGDKSIAIGDIAQ 29
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24
 XX
 RESULT 8
 ABU35023
 ID ABU35023 standard; protein; 1946 AA.
 XX
 XX ABU35023;
 XX
 XX 19-JUN-2003 (first entry)
 XX
 XX Protein encoded by Prokaryotic essential gene #20550.
 XX
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 XX Moraxella catarrhalis.
 XX
 XX WO200277183-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 23-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA38893.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 XX

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 62947; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1946 AA;

Query Match 44.1%; Score 86.5; DB 6; Length 1946;
 Best Local Similarity 50.0%; Pred. No. 0.053;
 Matches 20; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

QY 1 GIGISEA-DGGKGGANAR----GDKSIAIGDIAQAALGSQSIAG 39
 Db 48 GTAIQNNNSKGDRAETGNQSVAGKAVAGSSQAIAG 87

RESULT 9

AAW04505
 ID AAW04505 standard; protein; 1992 AA.

XX AC AAW04505;

DT 27-AUG-2003 (revised)
 DT 25-JAN-1997 (first entry)

XX Moraxella 200 kDa outer membrane protein.

XX Outer membrane protein; OMP; immunogen; vaccine; otitis media; diagnosis.

XX Moraxella catarrhalis.

XX WO9634960-A1.

PD 07-NOV-1996.

XX 29-APR-1996; 96WO-CA000264.

XX 01-MAY-1995; 95US-00431718.

XX 07-JUN-1995; 95US-00478370.

XX 26-MAR-1996; 96US-00621944.

XX (CONN-) CONNAUGHT LAB LTD.

XX Sasaki K, Harkness RE, Loosmore SM, Chong P, Klein MH;

XX WPI; 1996-506162/50.
 DR N-PSDB; AAT38740.

XX Moraxella outer membrane protein - useful as immunogen in protective
 PT vaccine and for diagnosis.

XX Claim 14; Fig 6; 109pp; English.

XX An approx. 200 kDa outer membrane protein (AAW04505) can be isolated from
 CC Moraxella catarrhalis otitis media strain 4223 by electroelution, or
 CC expressed from a gene (see also AAT38740) obtd. from a strain 4223
 CC genomic library. Natural or recombinant outer membrane protein is useful
 CC as an immunogen to protect against infection by Moraxella, esp. M.
 CC catarrhalis. It can also be used to detect antibodies, esp. for
 CC differential diagnosis between bacteria that cause similar symptoms, and
 CC also useful as a carrier for other antigens and used to raise antitumour
 CC antibodies for conjugation to therapeutic agents. (Updated on 27-AUG-2003
 CC to correct OS field.)

XX Sequence 1992 AA;

Query Match 42.9%; Score 84; DB 2; Length 1992;
 Best Local Similarity 56.4%; Pred. No. 0.11;
 Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR----GDKSIAIGDIAQAALGSQSIAG 39
 Db 1718 AKADGEAAVAIGRQTQAGNQSIAGDGAQATGDSIAIG 1756

RESULT 10

AAW69137

ID AAW69137 standard; protein; 1992 AA.

XX AC AAW69137;

DT 24-APR-2001 (first entry)

DE M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.

XX Moraxella catarrhalis strain Q8; major outer membrane protein;

XX 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 XX otitis media; detection.

XX Moraxella catarrhalis.

XX WO200107619-A1.

PD 01-FEB-2001.

XX 26-JUL-2000; 2000WO-CA000870.

XX 27-JUL-1999; 99US-00361619.

XX (CONN-) CONNAUGHT LAB LTD.

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX WPI; 2001-159722/16.

XX N-PSDB; AAF59106.

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis.

XX Claim 1; Fig 8A-V; 247pp; English.

XX The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as

CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the M.
 CC catarrhalis M56 200kDa protein in pKS348, which is given in the
 CC exemplification of the present invention
 XX
 XX
 SQ Sequence 1992 AA;

Query Match 42.9%; Score 84; DB 4; Length 1992;

Best Local Similarity 56.4%; Pred. No. 0.11; Indels 4; Gaps 1;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR---GDKSIAIGDIAQAALGSGSIAIG 39

Db 1718 AKADGEAAVAIGRTQAGNQSIAIGDVAQATGDSIAIG 1756

RESULT 11

AAB69133

ID AAB69133 standard; protein; 1992 AA.

AC AAB69133;

XX

XX 24-APR-2001 (first entry)

DT

DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.

XX

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection.

KW

XX Moraxella catarrhalis.

OS

XX WO200107619-A1.

XX

XX 01-FEB-2001.

XX

XX 26-JUL-2000; 2000WO-CA000870.

XX

XX 27-JUL-1999; 99US-00361619.

XX

XX (CONN-) CONNAUGHT LAB LTD.

XX

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX

XX WPI; 2001-159722/16.

XX

XX N-PSDB; AAF59100, AAF59101.

XX

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis.

XX

XX Example 3; Fig 2A-W; 247pp; English.

XX

XX The present invention describes an isolated and purified nucleic acid (I)

XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

XX The 200 kDa outer membrane protein (II) has antibacterial activity and

XX can be used in vaccines. (II), and its truncated versions, are used as

XX immunogenic compositions and vaccines to protect against M. catarrhalis

XX infections, particularly otitis media in humans. (II) is also used as

XX antigen in immunoassays for detecting specific antibodies (Ab), and to

XX generate Ab. (I) are used for recombinant production of (II) and its

XX fragments are used as probes for identifying/cloning 200 kDa protein

XX genes from other strains, and for diagnostic detection of M. catarrhalis.

XX (I) makes possible production of large amount of recombinant immunogens.

XX Expression of truncated versions of (II) reduces toxicity of the protein

XX towards the Escherichia coli host. The present sequence represents the M.

XX catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is used

CC in the exemplification of the present invention

XX

SQ Sequence 1992 AA;

Query Match 42.9%; Score 84; DB 4; Length 1992;

Best Local Similarity 56.4%; Pred. No. 0.11;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR---GDKSIAIGDIAQAALGSGSIAIG 39

Db 1718 AKADGEAAVAIGRTQAGNQSIAIGDVAQATGDSIAIG 1756

RESULT 12

AAB69134

ID AAB69134 standard; protein; 2047 AA.

XX

AC AAB69134;

XX

XX 24-APR-2001 (first entry)

XX

XX M. catarrhalis strain 4223 genomic 200kDa protein SEQ ID NO:7.

XX

XX Moraxella catarrhalis strain 4223; major outer membrane protein;

KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

KW otitis media; detection.

KW

XX Moraxella catarrhalis.

OS

XX WO200107619-A1.

XX

XX 01-FEB-2001.

XX

XX 26-JUL-2000; 2000WO-CA000870.

XX

XX 27-JUL-1999; 99US-00361619.

XX

XX (CONN-) CONNAUGHT LAB LTD.

XX

XX Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX

XX WPI; 2001-159722/16.

XX

XX N-PSDB; AAF59102, AAF59103.

XX

XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,

XX useful in protective vaccines and for diagnosis.

XX

XX Claim 1; Fig 3A-W; 247pp; English.

XX

XX The present invention describes an isolated and purified nucleic acid (I)

XX that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.

XX The 200 kDa outer membrane protein (II) has antibacterial activity and

XX can be used in vaccines. (II), and its truncated versions, are used as

XX immunogenic compositions and vaccines to protect against M. catarrhalis

XX infections, particularly otitis media in humans. (II) is also used as

XX antigen in immunoassays for detecting specific antibodies (Ab), and to

XX generate Ab. (I) are used for recombinant production of (II) and its

XX fragments are used as probes for identifying/cloning 200 kDa protein

XX genes from other strains, and for diagnostic detection of M. catarrhalis.

XX (I) makes possible production of large amount of recombinant immunogens.

XX Expression of truncated versions of (II) reduces toxicity of the protein

XX towards the Escherichia coli host. The present sequence represents the M.

XX catarrhalis strain 4223 genomic 200kDa protein, which is given in the

XX exemplification of the present invention

XX

SQ Sequence 2047 AA;

Query Match 42.9%; Score 84; DB 4; Length 2047;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR---GDKSIAIGDIAQAALGSGSIAIG 39

Db 1718 AKADGEAAVAIGRTQAGNQSIAIGDVAQATGDSIAIG 1756

Db 1773 AKADGEAAVAIGRTQAGNQSTAIGDQAQATGDSIAIG 1811

RESULT 13

AAB69135
ID AAB69135 standard; protein; 2053 AA.

XX AC AAB69135;

XX DT 24-APR-2001 (first entry)

XX DE M. catarrhalis strain Q8 200kDa protein SEQ ID NO:9.

XX KW Moraxella catarrhalis strain Q8; major outer membrane protein;

XX KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX KW otitis media; detection.

XX OS Moraxella catarrhalis.

XX PN WO200107619-A1.

XX PD 01-FEB-2001.

XX PF 26-JUL-2000; 2000WO-CA000870.

XX PR 27-JUL-1999; 99US-00361619.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Loosmore SM, Sasaki K, Yang Y, Klein MH;

XX DR WPI; 2001-159722/16.

XX DR N-PSDB; AAF59104.

XX PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
XX PT useful in protective vaccines and for diagnosis.
XX PS Claim 1; Fig 4A-V; 247pp; English.
XX CC The present invention describes an isolated and purified nucleic acid (I)
XX CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
XX CC The 200 kDa outer membrane protein (II) has antibacterial activity and
XX CC can be used in vaccines. (II), and its truncated versions, are used as
XX CC immunogenic compositions and vaccines to protect against M. catarrhalis
XX CC infections, particularly otitis media in humans. (II) is also used as
XX CC antigen in immunoassays for detecting specific antibodies (Ab), and to
XX CC generate Ab. (I) are used for recombinant production of (II) and its
XX CC fragments are used as probes for identifying/cloning 200 kDa protein
XX CC genes from other strains, and for diagnostic detection of M. catarrhalis.
XX CC (I) makes possible production of large amount of recombinant immunogens.
XX CC Expression of truncated versions of (II) reduces toxicity of the protein
XX CC towards the Escherichia coli host. The present sequence represents the M.
XX CC catarrhalis strain Q8 200kDa protein, which is given in the
XX CC exemplification of the present invention

XX SQ Sequence 2053 AA;

Query Match 42.9%; Score 84; DB 4; Length 2053;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR----GDKSTAIGDIAQALGQSQAIG 39

Db 1779 AKADGEAAVAIGRTQAGNQSTAIGDQAQATGDSIAIG 1817

RESULT 14

ABP71294
ID ABP71294 standard; protein; 2139 AA.

XX AC ABP71294;

XX DT 28-APR-2003 (first entry)

XX DE M. catarrhalis surface exposed IgD binding protein.

XX KW Surface exposed protein; bacterium; immunoglobulin; IgD; immunomodulator;
XX KW gene therapy; vaccine; mid.

XX OS Moraxella catarrhalis.

XX FH Key Location/Qualifiers

FT Peptide 1..66

FT Protein /note= "signal peptide"

FT Protein 67..2139

FT Region /note= "mature protein"

FT Region 764..913

FT Region /note= "specifically claimed immunogenic fragment"

FT Region 962..1200

FT Region /note= "specifically claimed immunogenic fragment"

PN WO2003004651-A1.

XX PD 16-JAN-2003.

XX PF 01-JUL-2002; 2002WO-SE001299.

XX PR 04-JUL-2001; 2001SE-00002410.

XX PA (FORS/) FORSGREN A.

XX PI Forsgren A, Riesbeck K, Janson H;

XX DR WPI; 2003-221598/21.

XX DR N-PSDB; ABZ58968.

XX PT New surface exposed immunoglobulin D-binding protein from Moraxella
XX PT catarrhalis, useful for treating an autoimmune disease or as vaccine,
XX PT comprises a molecular weight of 200 kDa.
XX PS Claim 1; Fig 6; 98pp; English.
XX CC The invention relates to a surface exposed immunoglobulin D-binding
XX CC protein detected in Moraxella catarrhalis. The protein has an apparent
XX CC molecular weight of 200 kDa, and can selectively bind membrane bound or
XX CC soluble IgD. The protein or its variant or fragment, is useful in
XX CC treating an autoimmune disease or as vaccine. The present sequence
XX CC represents a M. catarrhalis the surface exposed protein (mid) of the
XX CC invention

XX SQ Sequence 2139 AA;

Query Match 42.9%; Score 84; DB 6; Length 2139;

Best Local Similarity 56.4%; Pred. No. 0.12;

Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 5 SEADGGKGGANAR----GDKSTAIGDIAQALGQSQAIG 39

Db 1865 AKADGEAAVAIGRTQAGNQSTAIGDQAQATGDSIAIG 1903

RESULT 15

AAB69136

ID AAB69136 standard; protein; 2314 AA.

XX AC AAB69136;

XX DT 24-APR-2001 (first entry)

XX DE M. catarrhalis les1 200kDa protein SEQ ID NO:11.

XX KW Moraxella catarrhalis strain Q8; major outer membrane protein;

XX KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;

XX KW otitis media; detection.

XX OS Moraxella catarrhalis.

```
XX WO200107619-A1.
PN
XX
XX 01-FEB-2001.
PD
XX
XX 26-JUL-2000; 2000WO-CA000870.
XX PF
XX
XX 27-JUL-1999; 99US-00361619.
XX PR
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX PA
XX
XX Loomore SM, Sasaki K, Yang Y, Klein MH;
XX PI
XX
XX WPI; 2001-159722/16.
XX DR
XX N-PSDB; AAF59105.
XX DR
XX
XX New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
PT useful in protective vaccines and for diagnosis.
XX
XX Claim 1; Fig 5A-Y; 247pp; English.
XX PS
XX
XX The present invention describes an isolated and purified nucleic acid (I)
CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
CC The 200 kDa outer membrane protein (II) has antibacterial activity and
CC can be used in vaccines. (II), and its truncated versions, are used as
CC immunogenic compositions and vaccines to protect against M. catarrhalis
CC infections, particularly otitis media in humans. (II) is also used as
CC antigen in immunoassays for detecting specific antibodies (Ab), and to
CC generate Ab. (I) are used for recombinant production of (II) and its
CC fragments are used as probes for identifying/cloning 200 kDa protein
CC genes from other strains, and for diagnostic detection of M. catarrhalis.
CC (I) makes possible production of large amount of recombinant immunogens.
CC Expression of truncated versions of (II) reduces toxicity of the protein
CC towards the Escherichia coli host. The present sequence represents the M.
CC catarrhalis lesi 200kDa protein, which is given in the exemplification of
CC the present invention
XX
XX Sequence 2314 AA;
SQ
```

```
Query Match 42.9%; Score 84; DB 4; Length 2314;
Best Local Similarity 56.4%; Pred. No. 0.13;
Matches 22; Conservative 4; Mismatches 9; Indels 4; Gaps 1;
```

```
QY 5 SEADGGKGGANAR---GDKSIATGDIQAALGSGSIAIG 39
   :||| | | | | | | | | | | | | | | | | | |
Db 2040 AKADGEAAVAIGRTQTQAGNQSIAIGDNAQTGQSIAG 2078
```

Search completed: September 21, 2004, 23:40:28
Job time : 4.11791 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:36:00 ; Search time 215.568 Seconds
(without alignments)
3105.882 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 MNHIYKVFINKATGTFMAVA.....NGSADTQGHVGAAGAGPHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7316	68.3	1964	2	Q8KQM9	Q8kqm9 moraxella c
2	6475	60.5	2314	2	Q8KQM8	Q8kqm8 moraxella c
3	1246.5	11.6	2059	16	Q9PD50	Q9pd50 xylella fas
4	1233	11.5	2052	16	Q87DB1	Q87del xylella fas
5	1185.5	11.1	2353	2	P71401	P71401 haemophilus
6	1063	9.9	1778	16	Q8FCB2	Q8fcb2 escherichia
7	1039	9.7	2712	16	Q3F3X5	Q3f3x5 pasteurella
8	886	9.2	1588	16	Q8XDG4	Q8xdg4 escherichia
9	945	8.8	1004	2	Q8GM77	Q8gm77 haemophilus
10	939	8.8	1002	2	Q8GM78	Q8gm78 haemophilus
11	919.5	8.6	1461	16	Q8ZL64	Q8z164 salmonella
12	847.5	7.9	1315	16	Q87D62	Q87d62 xylella fas
13	831	7.8	1953	16	Q8BHJ2	Q8bhj2 rhizobium l
14	784.5	7.3	2351	16	Q8PCQ5	Q8pcq5 xanthomonas
15	763	7.1	1210	2	Q8GM75	Q8gm75 haemophilus
16	763	7.1	2190	16	Q8PGS0	Q8pgs0 xanthomonas

17	761	7.1	1204	2	Q8GM76	Q8gm76 haemophilus
18	756	7.1	1210	2	Q8GM74	Q8gm74 haemophilus
19	730.5	6.8	1299	16	Q9F3X6	Q9f3x6 pasteurella
20	728	6.8	1190	16	Q9PC04	Q9pc04 xylella fas
21	722	6.7	3930	16	Q98E20	Q98e20 rhizobium l
22	684	6.4	3420	16	Q8FUS1	Q8fus1 brucella su
23	682	6.4	1098	2	Q48152	Q48152 haemophilus
24	675.5	6.3	3692	2	Q48RR3	Q48rr3 fusobacteri
25	670	6.3	1096	2	Q8GM79	Q8gm79 haemophilus
26	670	6.3	1107	16	Q9F2D8	Q9f2d8 salmonella
27	670	6.3	1116	16	Q83ST9	Q83st9 salmonella
28	669	6.2	2201	2	Q8GF46	Q8gf46 zymomonas m
29	669	6.2	3705	16	Q8ZHA1	Q8zhal yersinia pe
30	669	6.2	3710	16	Q8CZU2	Q8czu2 yersinia pe
31	658.5	6.1	3165	16	Q8RDQ9	Q8rdq9 fusobacteri
32	654.5	6.1	3705	2	Q9F285	Q9f285 yersinia pe
33	653	6.1	3133	2	Q8KER1	Q8krr1 fusobacteri
34	651.5	6.1	2340	16	Q9ZD91	Q9zd91 rickettsia
35	650.5	6.1	6274	16	Q880E1	Q880el pseudomonas
36	646	6.0	1264	2	Q8RQ61	Q8rd61 actinobacill
37	645.5	6.0	5291	16	Q8X2T1	Q8x2t1 escherichia
38	643.5	6.0	2411	16	Q8PF72	Q8pf72 xanthomonas
39	636.5	5.9	3961	16	Q8BP94	Q8bp94 xanthomonas
40	635.5	5.9	990	16	Q83J22	Q83j22 shigella fl
41	634	5.9	5188	16	Q8X4H5	Q8x4h5 escherichia
42	630.5	5.9	3322	16	Q8XQZ5	Q8xqz5 raistonia s
43	628	5.9	3441	16	Q89PB9	Q89pb9 bradyrhizob
44	624	5.8	1309	16	Q8XPL8	Q8xpl8 raistonia s
45	623.5	5.8	2806	16	Q8RI19	Q8ril9 fusobacteri

ALIGNMENTS

RESULT 1

Q8KQM9 PRELIMINARY; PRT; 1964 AA.
 ID Q8KQM9
 AC Q8KQM9
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Hemagglutinin.
 GN HAG
 OS Moraxella catarrhalis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Moraxellaceae; Moraxella.
 OX NCBI_TaxID=480;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O35E;
 RX MEDLINE=22112901; PubMed=12117964;
 RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Gene J.W. III,
 RA Hansen E.J.;
 RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
 RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
 RT Activities.";
 RL Infect. Immun. 70:4523-4533 (2002).
 DR EMBL; AY077637; AAL78284.1; -.
 DR InterPro; IPR008640; Hep Hag.
 DR Q9pd50 xylella fas
 DR InterPro; IPR008635; HIM_
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep Hag; 9.
 DR Pfam; PF05662; HIM; 7.
 DR Pfam; PF03895; Yada; 1.
 SQ SEQUENCE 1964 AA; 201585 MW; 764A08F5F1F6854E CRC64;

Query Match

Best Local Similarity 68.3%; Score 7316; DB 2; Length 1964;

Matches 1541; Conservative 127; Mismatches 209; Indels 332; Gaps 29;

QY 1 MNHIYKVFINKATGTFMAVAAYAKSHSTGGSCATGQGVSVRTLSFARIALAVLIGAT 60

DB 1 MNHIYKVFINKATGTFMAVAAYAKSHSTGGSCATGQGVSVRTLSFARIALAVLIGAT 59

Qy	61	LANGSAYA-GTIGISEADGGKGGANARGDKSIAIGDIAQALGSSOIAIGDNKI VHNSNNAN	119
Db	60	LANGSAYAQGIAIGSGGELNGGARADGEKSIAIGSIAQATGSSOIAIGDNKAVSNSNNAN	119
Qy	120	IGAKASNESIAITGGDVLAGSHASIAIGSDLLYLKKEITVQQ--ISELLPIIRGOKALNDIY	178
Db	120	IGARASGESEAIGGVDVAVSGNASIAIGSDLLYLNHDSKNGTNTQNLISQHRILKNT--	178
Qy	179	QLADTNLQVRRTHAQCHASTAVGMSYAKGFHSNAGFTGATAGTYSYSLAVGLTATAKAA	238
Db	179	--RDXN--QYRRTTASGHASTAVGMSHATGFHANAFGTYSTASGNVALALGLASSASGG	234
Qy	239	SSIAVGSNAQAI GFAATAVGGSTQVNLNRGIALGFGSOVLQKDNVDVNAANVRAYAP--D	295
Db	235	TSVAIGSNAEAKYV GAGALGSDTQVDLNYGIALGYGSOILNNNN--NNNAYMFKATD	290
Qy	296	DNOPIDNRYATPFNGATDVFSIGNSNGNDSIBRKIINVAGSADTDVNAVQLEAVPL	355
Db	291	QNFISQSSHKAS---TWGPLSIGNS---TQKRIINVAGSQDITDVAVQLEAARVAL	342
Qy	356	ANFOITFKGDDSNRVEKGLGKLTITGGAQ--TSALTD--HNI GVVO--NGDGLKVQLAETL	412
Db	343	ANRKIITFGSGGN--VQKGLNETLITKGGETIDNKLTGDNIGVVKDNNSLGVKLAKTL	401
Qy	413	TSLKQVVTENLTANEKVTVG-----	432
Db	402	TGLETVNTQNLATDTITVANNTAKLLNSGLTFSSNPPTNTSKTVYIDGVKFTDNQ	461
Qy	433	-----KTRLTDXIGFTNDMNGIDESKPYLD-----KDTGTHAGGOKITKLT	474
Db	462	NAAIPDTRITNEKIGFAKADGTVDQAPSLDKKQLKGVSVTITKDDGINAGKQKISNVK	521
Qy	475	AGVVDDAAATVQOLKKNQTAESALQFTVKKYVDKNGNDANDSKIIITVGKNNKPDGTQVN	534
Db	522	EATDDTDVYIKHAKVQTADSALOFIQVQ-----DNDDINATVAK--ETNGKFN	573
Qy	535	TLKLKGVGNDV--TTEINGVTIFGLNQNGGLTVGNSTLNNDGLSVKNTNSNKIOI QVGADG	593
Db	574	ILKLKGVNLNVKNTGTGVTFSIKQDGLDITGKSTLNNNGLTVK--NDNEIOIQVSADG	631
Qy	594	ITPTDISNSKPGAGIENTRITRIGI GIPANNTGSLDANKPRLPTGINAGGKELTNVQSA	653
Db	632	IKFTGIGNGLGTSIENTRITKONI GFAONTGSLDTNKPHLT-----	674
Qy	654	INPATNGQLDFMNRSLSTANTEKSGSAATIKDLYNLSQVPLTFAGDTGPNVTKLGEILK	713
Db	675	-----	674
Qy	714	VKGKTTADDLTKNNIGVWADSTONSUTVKLAKTLSDLDVNTKTLTASDKVTVDGSNNT	773
Db	675	-----	674
Qy	774	AKLQNGDLTFQKQNTGATPATNSKTI GVDGLKFTDNNGIALDGTITYITKDKGVFAKQDGS	833
Db	675	-----	674
Qy	834	LDKSPYLKDKLKYGEVEITTINGINAGGKAITGLSNTLTATNATTHGVTLQGLVDSTD	893
Db	675	-----KDRLKVDGVEITSTGINAGSKAITGLSSTLPSIASQSQRNIELGNTIEDKD	725
Qy	894	KTRAASIGDVLNAGPNLKNNGDAKDFVSTYDVTDFINGNATTAKVTYD--GKASKVAYDV	951
Db	726	KXNAASI SDVLNAGNLQNGIAVDVSTYDVTVPFINGNATTAKVTYDEANKTSKVAIDV	785
Qy	952	NVDGTTIHLTGADGNKNOIGVKTITTLTKIDAKGDKAINFSVNSGDDKALINAKDIALNLN	1011
Db	786	NVDNKTIELTGDNGTNKIGVKTITTHSTNGAT--SFNAGDDNALVNAKGIADNLN	842
Qy	1012	TLAGEIRNTKGTADTALQTFQYKVKVKNGGDDDDADTITVQKDA--KTNQVNTLKLKGN	1069
Db	843	TLAKEIHUTKGTADTALQTFKVK-----DGATDDETTITKGTQGTQNGTNTLTLKGN	897

Db 1369 WSPKPKPQMGLTQAMNVKSVINKEQVNDANKQGINEDNAFVKGLEKAAADNKTGNAVT 1428
Qy 1563 VGLNVAQPTPLTFAGDTGTAKLGETLTIKGGQTDNKLTDNNIGVAGTDTGFTVKLA 1622
Db 1429 VGLNVAQPTPLTFAGDTGTAKLGETLTIKGGQTDNKLTDNNIGVAGTDTGFTVKLA 1488
Qy 1623 KDLTNLNSVNAAGTKIDDKGVSVFVDSGQAKANTPVLSSANGLDL- 1666
Db 1489 KDLTNLNSVNAAGTKIDDKGVSVFVDSGQAKANTPVLSSANGLDLGGKREISNIGNAVDND 1548
Qy 1667 - 1666
Db 1549 AVNFKQNEVAKTVNVLNNSGASLPFVVTDANGKPIGTGDKPKQAIKAGADKYHA 1608
Qy 1667 - 1666
Db 1609 NANGVPVDKGNPITDADKLANLAVHGKPLDAGHVVASLGGNSDAITLTIKSTLPDIA 1668
Qy 1667 - 1666
Db 1669 TPNTGNANAGAAQSLPSLRAAQSSNAASVKDVLNVGFNLQTNHENVDFVKAYDTVNFVG 1728
Qy 1667 - 1666
Db 1729 TGADITSVRSAGDTMSNITVNTALATDDGCVNLKAKDGKFKYKADLMPGSLKAGKSA 1788
Qy 1667 - 1666
Db 1789 SDAKPTGLSLVNPNAKSGTGDVAVALNLSKAVFKSKOGTTTTTSSDGSISQKDNSS 1848
Qy 1667 - 1666
Db 1849 ITLSKGLNVGKGVISNVGKGTDTDAANVQQLNEVNLGLGNAGNDNADGQVNIADI 1908
Qy 1717 KDPNSGSSSNRTVTKAGTVLGGKNDTEKLATGGVQGVVDKGNANGDLNNVVKTKQ 1776
Db 1909 KDPNSGSSSNRTVTKAGTVLGGKNDTEKLATGGVQGVVDKGNANGDLNNVVKTKQ 1968
Qy 1777 DGSKALLATYNAAGQNTLNNPABAI DR INEQIRFFHNDGNQBPVQVGRNGIDSSA 1836
Db 1969 DGSKALLATYNAAGQNTLNNPABAI DR INEQIRFFHNDGNQBPVQVGRNGIDSSA 2028
Qy 1837 SGKHSVAIGFOAKADGAEAAVIGRQTAGNQSTAGIDNAQATGDSQSTAGTGNVAGKHS 1896
Db 2029 SGKHSVAIGFOAKADGAEAAVIGRQTAGNQSTAGIDNAQATGDSQSTAGTGNVAGKHS 2088
Qy 1897 GATGDPSTVKADNSYSYVGNNOFTDATQTDVFGVGNNTVTSNSVALGNSAISAGTHA 1956
Db 2089 GATGDPSTVKADNSYSYVGNNOFTDATQTDVFGVGNNTVTSNSVALGNSAISAGTHA 2148
Qy 1957 GTQAKKSDGTAGTTTATAGTGVKGFAGQTAGVAVSVGASGAERRIQNVAAGEVSATSTD 2016
Db 2149 GTQAKKSDGTAGTTTATAGTGVKGFAGQTAGVAVSVGASGAERRIQNVAAGEVSATSTD 2208
Qy 2017 AVNGSOLYKATQCIANATNLDHRIHQENKANAGISSAMAMASMPQAYIPGRSMVTGGI 2076
Db 2209 AVNGSOLYKATQCIANATNLDHRIHQENKANAGISSAMAMASMPQAYIPGRSMVTGGI 2268
Qy 2077 ATHNGQAVAGLGLSKLSDNGQWPFKINGSADTQGHVCAAVGAGHF 2122
Db 2269 ATHNGQAVAGLGLSKLSDNGQWPFKINGSADTQGHVCAAVGAGHF 2314

RESULT 3

Q9PD50 PRELIMINARY; PRT; 2059 AA.
ID Q9PD50
AC Q9PD50;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Surface protein.
GN XF1529.
OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OX Xanthomonadaceae; Xylella.
RN NCBI_TaxID=2371,
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=9a5C;
RZ MEDLINE=20365717; PubMed=10913047;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.W., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garner R., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuranai E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto R.B., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.P., Truffi D., Teai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RL "The genome sequence of the plant pathogen Xylella fastidiosa";
RL Nature 406:151-159(2000).
DR EMBL; AE003982; AAF84338.1; -.
DR PIR; D82671; D82671.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM-
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 6.
DR Pfam; PF05662; Hep Hag; 15.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2059 AA; 204035 MW; EBA6A71B5DD24E10 CRC64;
Query Match 11.6%; Score 1246.5; DB 16; Length 2059;
Best Local Similarity 24.3%; Pred. No. 8.6e-32;
Matches 578; Conservative 344; Mismatches 780; Indels 681; Gaps 108;
Qy 62 NGSAYAGIGISEADGG---KGGANARGDKSTAI-----GDIAQA 97
Db 36 NAQVYINDGM---DGGQRIVDNGSAGSVERTVATQCSDEDPWTVTGVARFFSGSTAAEQ 92
Qy 98 LG-SQSIAGIDNKIVHNSNNANIGAKSGNESIAIG-GDVL---ASGHASIAIGSDDL 151
Db 93 QGASRNLTIGGSLV---NSGQGVNDVLNKYSIRMGSVITMNTVAGTNAIAIGS--- 145
Qy 152 YLKKETVQIISELLPIIRGQKALNDIYQLADTNLQKVRTHAQGHASTAVCA--MSYAKG 209
Db 146 -----AQSSAADALKASLATKASG 164
Qy 210 HFSNAGFTRATAGTYSLAVGLTATA-KAASSIAGSVNAQAIQFAATAVGGSTQVNLNRG 268
Db 165 ARAIAIGAKASADGVDTVALGSGATAGTGASSIAGLNASAVN-GAVAVGGGALVTPDG 223
Qy 269 -IALGFGSQVLQKNDVNAANVAYAPDDNQPIDNRYKATFKNGATDVFSIGNSGNDSI 327
Db 224 AVALGLNSVA-----STGKGLSGYDP-KTKTSTDSAAWKSTLAHV-SIGDVSTNLK 275
Qy 328 RRKIINVGASADTDVNVNAQLKEAVRLNR--QITFKGDDSNRVE-----KGLGKTL 379
Db 276 TROLGLAGTSNTDAVNVAQLKVDIEASRGWNLTASGANSNVAPGSSVDLKNLTKD 335

GN HSF OR PD0744.
OS *Xylella fastidiosa* (strain Temeculal / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
RN NCBI_TaxID=183190;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2241331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Camarvan F.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of *Xylella*
RT *fastidiosa*,"
RL J. Bacteriol. 185:1018-1026(2003).
DR EMBL: AE012556; AAC28613.1; -
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 4.
DR Pfam; PF05662; HIM; 17.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2052 AA; 203200 MW; CF194A71FF197655 CRC64;

Query Match 11.5%; Score 1233; DB 16; Length 2052;
Best Local Similarity 23.9%; Pred. No. 2.3e-31;
Matches 583; Conservative 345; Mismatches 779; Indels 730; Gaps 113;

QY 23 AKSHSTGGSCATGCVGVSRTLSFARIALAVLIGATLNGSAVAGI--GISEADG--- 76
DB 9 AGTYSGNLKSVDVEAAGAIHLQADSPKFGNVIN---NGKISGLTAGTDTDAVNL 65
QY 77 -----GKG-----GANARGKSIAG--DIAQALGSOSIA-IGD-NKIVHNSNNAN 119
DB 66 QLKXISDVTDKGWTLTASGANGSKVSGGAVDLAKNTDGNLTISKSGSDNVVFNLSKDFK 125
QY 120 IGAKASG-----NESIAGGDV-----LASGHASIAIGSD--DIYLKKTQQISEL 164
DB 126 VDGMTSGTTVVNDGVRGVTDAVLTGLTIANGPAVTSAGIDAGSKVISHVAAGEVSET 185
QY 165 LPIIRGQKALNDIVQLADT-----NLQVRRTHAQ-----GHASTAVGAMSYAKGHFSN 213
DB 186 STDVAVGSQLNAVQAASQPVTFSGNEGSKVRLTGQAVVIGESST---AGTYSGNLKS 242
QY 214 AFGTRATAEGTYSILAV-----GLTATAKAASIIAVG-----SNAQAI 250
DB 243 VVD---EAAGAIHLQADSPKFGNVINNGKISGLTAGTDTDAVNLSQLKSISDVTDK 299
QY 251 GFAATAVGGSTQVNLNGIALGFQSGVQLQKNDVNAANVRAYAPDDNQPDNRVK----- 305
DB 300 GWTLTASGAN-----GSKVSGGAVDLAKNTDGNLTISK--SGSDNVVFNLSKDFKVD 350
QY 306 -----ATFKNG-----ATDV-----PSIGN-----SNGNDSIRKLIINVAG--SADT 341
DB 351 GMTSGTTVVNDGVRGVTDAVLTGLTIANGPAVTSAGIDAGSKVISHVAAGEVSET 410
QY 342 DAVNVAQLKEAVRLANRQITPKGDDSNRVEKGLKTLITIGGAQTSALT DHTNIGVVQNG 401
DB 411 DAVNGSQLNAVQAASQPVTFSGNEGS--VKRTLQAVVIGESSTAG-----TVSG 460
QY 402 DGLKVQLAETLTSLKMTTENLTANKEVTVGKLTLLTDKIGFTNDMNGIDESKPYLDKDT 461

DB 461 GNLSKVDEAAGAIHLQAD-----SPKFG-----NV 487
QY 462 GIHAGGOKITKLTAGVVDDDAATVYGLKVNQTAESALQTFVTKVDK-----NGNDANDS 517
DB 488 VINNGG-KISGLTAGTDTDAVNLSQLKSISDT-----VDKGWTLTASGANGS 534
QY 518 KIITVGKNNKPDGTQVNTLKLKGEGVDVTTETNG---TVTFLGNQN---NGLTVGNSTL 571
DB 535 KVVSGG-----AVDLKNTDG-NLTISKSGSDNVVFNLSKDFKVDGMTSGTTVV 582
QY 572 NNDGLSVNTNSNKQIQVGADGITFTDINSKFGAGIENTRITRDGIGFANNTGSLDAN 631
DB 583 NNDGVKV-----GSDVALGTTGLTIT-----N 604
QY 632 KPLRTPGINAGKELTN-----VQSAINPATNGGQLDFMRLSTANTKSGSAATIKDL 686
DB 605 GPVATSGIDAGSKVISHVAAGEVSETSDAVNGSQLNAVQAASQ----- 650
QY 687 YNL-SQVPLTFAGDTGPNVTKKLGEILKVKGGKTTADDLTKNNIGVAVDSTNSLTVKLAK 746
DB 651 -----PVTFSGNEG-SVKRTLQAVVIGESSTAGTYSGNLKSVDVEAAGAIHLQAD 703
QY 747 T-----LSDLDVNTK-----TLTAS-----DKVTVDS 769
DB 704 SPKFGNVINNGKISGLTAGTDTDAVNLSQLKSISDVTDKGWTLTASGANGSKV-VSG 762
QY 770 GNNTAKLQNGDLTFSKQNTGATPATN-SKTIGVDGLKFTDNGGIALDGTYYITKD--KVG 826
DB 763 GAVDLKNTDGNLTISKSGSDNVVFNLSKDFKVDGM-----TSGTTVNNNDGVKVG 813
QY 827 FAKQDGSLDKSPYLDKDKLVQGE-VEITNGINAGKAIT-----GLSNTLTDATNATT 880
DB 814 -----TDVALGTTGLTIANGPAVTSAGIDAGSKVISHVAAGEVSETSDAVNG-- 861
QY 881 GHVTOLGIVDSTDKTRAASIGDVLNAGFNKNNGDAKDFVSTYDVTDFINGNATTAQVY 940
DB 862 ---SQLNAV-----QAQASQPVTFSG-----NEGSVK---RTLQAVVIGESSTAG-TY 904
QY 941 DGKASKVAYDVNVDDGTTIHLTGADGNKNQIGVKTTLTKTDAKGDKAIFSVNSGDDKAL 1000
DB 905 SGNLKSVDVEAAG--AIHLQADSPK--FG-----NVLINNG----- 938
QY 1001 INAKDIADNLNTLAGEIRN-TKGTAOT-ALQTFQVKVKGNGDDDDADITVKGDAKTN 1058
DB 939 -----GKISGLTAGTDTDAVNLSQLKSI-----SDVTDKGWTLTAS 975
QY 1059 QVNTLKLKNGGLDIQTNKDGTVTFGINTOS-----GLKAGNNTLNNL 1104
DB 976 GANGSKVSGGAVDLK-NTDGNLTISKSGSDNVVFNLSKDFKVDGMTSG-TTVVNDG-- 1032
QY 1105 SIKNTAGNEIQVGADGVKFAKVNNGVVGAGIDGTTTIRTRDEIGFAGTNGSLDKSKPHLS 1164
DB 1033 -----VKVGS-----VALGTTGLTI-----TNG-----PAVT 1055
QY 1165 KGINAGKKITINIOSGEIAQNSNDVATGKIYDLKTELKNSISSTAKTAQNSLHFSVA 1224
DB 1056 ASGIDAGSKVISHVAAGEVSETSDAVNGSQL-----NAVOAQA----- 1094
QY 1225 DEQNNFTVSNPYSSYDTSKTSVDITFAGENG-IITTKVNGVVRVGIDQTKGLTTPKLT 1283
DB 1095 -----SQPVTFSGNEGSKVRLTGQAVVIGESSTAG-----TY 1127
QY 1284 GNNNGKIVIDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGI-----IKDEDK--T 1334
DB 1128 SGNLKSVDVEAAGAIHLQADSPKFGNVLINNG-----GKISGLTAGTDTDAVNL 1180
QY 1335 RAASIVDLVLSAGFNLQNGE-----AVDFVSTVDTVNFADGNATTAQVTDYDTSKT 1385
DB 1181 QLKXISDVTDKGWTLTASGANGSKVSGGAVDLKNT-----DGNLTISK-----SGDS 1258
QY 1386 SKVYDVNVND-----DVTIEVKDKKLGKVTTLTSTGTGANKFALSNOATGDALVKA 1437
DB 1229 NDVVFNLSKDFKVDGMTSGTTVVND---GVKVGSDVALGTTGLTIANGPAVTSAGIDAG 1285


```

Db 1506 AQVSQNTQSTINLTQVT-----NLDTR-----VTNLENGIGDIV----- 1540
QY 1774 TQDGGKXKALLATYNAAQNTYLTNNPAEALDRINEQIRFFHVNDGNQBPVVQGRGID 1333
Db 1541 -----TTGSTKYFTN-----TDGAD 1556
QY 1834 SSASGKHSVAIFQAKADGAAVAIGRQTQAGNCSIAIGDNAQATQDSIAIGTGNVAVG 1893
Db 1557 ANAQKDSVAI-----GSGSIAADN-----SVALTGGSV--- 1586
QY 1894 KHSAGIDPSTVKADNSYSYGNMNFDTATQTDVFGVGNMITYTESNSVA-LGSNSAISA 1952
Db 1587 -----ADEENTISVGS-----TQRRITNVAAGVATDAVNVSLKSSEA--G 1628
QY 1953 GTIAGTQAKSDGTAGTTTTAGATGTVKGAGQTAAGVAVSGASGABRRIONVAAGEVSA 2012
Db 1629 GVRDYTKADGSDIDYSNITLGGNSGTT-----RISNVASG---V 1664
QY 2013 TSTDVANGSOLYKATQCIANATN-----ELDRHIONENKANAGISSAMAMASMPQAYIPG 2068
Db 1665 NNNDVNNYAQLKQSVQFTKQYTDQRMVEMDNKLSKTESKLSGGIASAMATGLPQAYTPG 1724
QY 2069 RSMVTGGIATHNGGAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAGVAGFHF 2122
Db 1725 ASMASIGGGTYNGESAVALGVSMYANGRWYKLGSTNSQGEYSAAALGAGIQW 1778

RESULT 7
Q9F3X5 PRELIMINARY; PRT; 2712 AA.
AC Q9F3X5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MapB protein (Hsf).
GN MAPB OR Hsf 1 OR PM0714.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RL "Complete genomic sequence of Pasteurella multocida Pm70.";
DR EMBL; AJ277636; CAC14203.1;
DR EMBL; AE006108; AAK02798.1;
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; YadaA.
DR Pfam; PF05658; Hep Hag; 14.
DR Pfam; PF05662; HIM; 10.
DR Pfam; PF03895; YadaA; 1.
KW Complete proteome.
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;

```

```

Query Match 9.7%; Score 1039; DB 16; Length 2712;
Best Local Similarity 23.7%; Pred. No. 4.7e-25;
Matches 618; Conservative 326; Mismatches 888; Indels 778; Gaps 127;

QY 12 ATGTTFMAVAEYAKSHSTGGSCATGQVGSVRTLTFARIAALAVLVIGATLNGSAVAGIGI 71
Db 374 ASDNSIALGNARTNKDSSIAIG-----DSAETKAAHSLAVGTTSKALAEALAL 425

```

```

QY 72 SEADGGKGGANARFGDKSIAIGDIAQALGSSOIAIGNKIVHNSNNNANI GAKASGN--ES 129
Db 426 GKL-----AEAKGTSVAMGNTSKADGNSVAVGNTSQTLSQNTIA-IGSSAIANPERT 478
QY 130 IATG-----CDVLASGHASIAIGSDDDLYLKKEVTQOISELLPIIRGOKALND---I 177
Db 479 ISIGLNAGKQGEADATGTTKHSQINIGENS---GEGVIGQLNIGIGHAGTNVVGKHNIAL 535
QY 178 YQLADTNLQKRYRTHAQ-----GH-----ASTAVGAMS YAKGHFSNAGFTRATA 221
Db 536 GTVAGTNLKNSETSKANVISIGHEANKYDOLTAVQKSTVVGAQTKAASR-STALGAATA 594
QY 222 EGYSLAVGLTATATAKAASSIAVGSNAQ-----AIGFAATAV----- 257
Db 595 LGLDAVAVGITSKAEGBDKSVAIGANSTADSNVALGATSRVAKEGSGYLTKQSSILVVS 654
QY 258 -----GGSTQVNLNRGIALGFG-----SOVLQK-DNDVNAA-NVRVAPDQ 298
Db 655 VHQDGAADQHILRLRVNADGVDEQDAATVAQLKKVTEKVTSDLOQFALSHAPTASE 714
QY 299 PIDNRYKATFKNGATDVFSIGNSNGNDSIRRKIINVGAGSADTDVAVNAQLKEAVRLANR 358
Db 715 IKDYTPPPPTGAENKITL-----KAKTRISNVAPAEALD TDVNLGVNVHIV-TKNK 766
QY 359 QITFKGDS-----NNRVEKGLGKLTITFGGACTSALTDHNIIGVVON----- 400
Db 767 AHYFSVNDTGINPVPGNHNDGATAKLAMAIGAQNAAKAERSVAIGNNTTVNGEGSIGLG 826
QY 401 -----CDGLKVQLAETLT-----SLKMVTTENLTANEKVTVGKLTETDKI 441
Db 827 TYKGSALDDGVPRESEITVTKPSKNVKYIGIAGAGTTTD--GNNSIALGSLAATSDK- 883
QY 442 GFTNDMNGIDESKPYLDKDTGIHAGGQKITKLTAGVVDDDAATYVQLK-----KVNQTA 495
Db 884 -----NPGANVDRAIALGYNAVSSAEKANAI GDRAVANSVKGNAGFSQALSGA 931
QY 496 ESALQTTTVKVKDKNG-----NDANSKIITVGNKNKPDGTQVNTLKLKKGENDVTT 548
Db 932 ESSTAIGTESKSEQNAYALGTSHAKGLNSIAFGTINQVSVGQNSGSIYAGELGNAKAT 991
QY 549 ETWGTVTFGI-NQNNGLTVGNSTL--NNDGLSV-----KNT----- 581
Db 992 VINGEGTSLGNTNSTLTANESGIFGNSNEIKAKENARI VGNKNNTIGAI BEKPHVGP 1051
QY 582 -----NSNKQIQVAGDGI-----TFTDIS- 600
Db 1052 AAPVNDLKDIIYVTGDNKISSDKKLAKDLSGLFVYGHNNIAQLPDPDSTEEFTLTDSV 1111
QY 601 -----NSK-----PGAGIENTTTRID-GIGFANNUGSLDA-----NKPR 634
Db 1112 IGANNLTNTKGNFYVLGNVNTATLNSVYLGAADSAYTTGNTSTSMNSMADMANGLNKS 1171
QY 635 LT-----PTG---INAGGKE--LTNVOS-----AINPATNGGOLDFMNRLLSTAN 679
Db 1172 YTFAGSQFVGVTVGAKERRVQNVASGLVTEASTDAINGSQLFALTR----- 1220
QY 680 AATIKDLNLSQVPLTFAGD---TGNP-----VTKKLGBILKYVGKTKTADDLT- 725
Db 1221 -----PLRFAGDNSTLSNPNKPGTDTVTISRSSNQGMKVGVGENDGNKLT 1267
QY 726 --KNNIGVADSTNSLTVKLAKTLSDI-DAV-----NTKILTASDKVTVDSGNNTAKLON 778
Db 1268 IADKNIGVIANG-DHTLEVRILAKTLNLKQATFGTDTKTINKDGMHTTNGANTVSLTE 1326
QY 779 GDLTFSKQNTGATPATN-----SKTIGVDGLKFTDNNIGIAL---DGTITVITDKVGF 828
Db 1327 GGL-----NNGKNKITVVAAGQNETDAVNVKQLNDLKAEGFLTGEGDQT--VKQALG 1379
QY 829 -KODGSLD-KSKPYLDKD---KLKVG-EVEITTINGINAGG-----KAITGLSNTL 872
Db 1380 IKVTGDDNVKTKIVTDADGSKKLEIGLENQVTLGGEAKNGNPAADGKLTNLKQAGTKV 1439
QY 873 TDATNATTGHTQLG-----IVDSTDKTR-----AASIGD 902

```

Db 1440 LOGANGTVGLTGADGAQAVITVKGRPTLDNAETPRIAGNEEVATLNDGLKFGNAGD 1499
QY 903 VLNAGPNLKNNGDAKDFVSTYDITD-----FINGNATT-----AKVTYDGKAS 945
Db 1500 VHNAKLNTQVDVKGATANTVWDFKQGNIMTRVEGNTITVALAKALSGLTSATFGDPAS 1559
QY 946 --KVAYDVNVVG-----TTIHLT--GADGNKNOI-----GVKTTTLTKDAKDKALN 989
Db 1560 NPKDSTVINKOGLTTTQGDNTVSLTDDGLDNGKQIKNVASGLTTTNGTATSLDDAVOT 1619
QY 990 FSVNSGDDKALINAKDIADNLNLTAGEIRNTKGTADTALTQTFQVKKVKGNGDDNDADTI 1049
Db 1620 NGVNVGDLKALN--NITNGTNPLOG-----FGLKDKAGNTFK-----QNLGETA 1662
QY 1050 TVGKDAKNTQVNTLKLKGN--GLDITQNKDGTVTFTGINTQSGLKAGNNTLNNGLSILKN 1108
Db 1663 QITGD---SNVNTKVVDGQGGKALEVSLANQLTLKGPEANVP---NATGAGKITLKD 1716
QY 1109 TAGNEQIQV--GADG-----VVKFKNVNVGVGAGIDGT-----TR 1141
Db 1717 DKGTRVVVDGEGALSITGQATQGAAPTAKIKVAGNPDLENTSDDPANNQNKTR 1776
QY 1142 ITRDEIGFAGTNGSLDKSPHLSKDGINAGGKKITNIQSGEIAQNSNDAVTGGKIYDLKT 1201
Db 1777 ITYD---IAGPNGTIVTEQLATLNDGLKFGA-----NTGDV---HDAKLNTRV-DVKG 1822
QY 1202 ELENKISSTAKTAQNSLHFEFSVADSGGNFTVS-----NPSYSSYDSTKSD 1247
Db 1823 KAENTNWFADAGQNTMTQIS-----GNTITVALAKALAGLDSATFGNP---ADGSKDGA 1874
QY 1248 VITFAGENGITTKVANKGVVRVGDIDTKGLTTPKLT-VGNNGKGVVIDSQNGNITIGLS 1306
Db 1875 VI-----NND-----GLTTTEGDKTVKLTTEKGLDNGGNQINLNDVSLGKKTGDG-- 1916
QY 1307 NTLANVTNDKGSVRTTEQGIKIDEDKTRAASIVDVLSA---GFNLQNGREAVDFVSTYD 1363
Db 1917 -SVVALKDAEGSVLT--NGVNVGDLKN---AIKDVTSATNGSGFLKKKAGA----- 1961
QY 1364 TVNFDAGNATTAKVYVDTSKTSKVYVNVDDTIEVKDKLGVKTTTLTSTGTGANKF 1423
Db 1962 --EFKQDLGTTTAQITGDKNINT-KVIDVPNSNDKALEI----- 1996
QY 1424 ALSNQATGDALVKASDIVAHLNLTSGDIQTAK-GASQANSSAGYVDADGNKVIYDSTDNK 1482
Db 1997 -----SLANDITLGRKAGDGVDSGLVNGKOGASVVLNG----- 2030
QY 1483 YYQAKNDGTVDRTKEVAKDKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANKKQGINEQNA 1542
Db 2031 -----KDGSIGLTGRQD-----GSDGKKSATISVK-----DGRAGVDGKD- 2066
QY 1543 FVKGLEKAASDNKTNA-AVTVDGLNVAQTPLTFTAGDTG-TTAKLGETLTIKGG----- 1596
Db 2067 ---GDTKTRIVYETKDATKPVVEEATLNDGMKFVNGDKREVTRKLNETLIDKGLDAA 2123
QY 1597 -QTDNKLTDNNIG-----VVAGTDG--FTVKLAKD----- 1624
Db 2124 TVADNAKVSSSLGVLKNTNAEGTGLEIVMKERTFFSGLVNVNGKDGEDAAVKFARKDGKMS 2183
QY 1625 ---LTLNLSVNAAGGTIKIDK---GVSFVDSGQAKANTPVLSSANGLDGGKVIQNVGKG 1677
Db 2184 IAAVTD-NDGNATGLTIKDKGNPGVTF--NNDR-----ITNVTAG 2222
QY 1678 TKDTPAANVQVLEVRNLLGLGNAGNDNADGNQVNI-A-DIKKDPNSGSS-----N 1727
Db 2223 VDDKDAVNVSQLD-----GLAKATTKVEAGKNMTVPTVNQD---GSTTYTVATEDNVN 2274
QY 1728 RTVIRAG-TVLGGKGNNDTEK-----LATGGVQGVGDK-DGNANGDL 1767
Db 2275 FTTVTGTNTW-----NNDGVKVGDNVNLNEGKAGDVTVTTAGINAGNKKYTGVDAGDI 2330
QY 1768 S-----NVWVKTKQKGSKKALLATYNAAGQTNVLTNNPAAIDRINEQG-IRFFHVN 1818

Db 2331 SPNSTDAVNGSQLNAVKETABAGWHLTANGADSSNV---KPRNTVDLNNTDGNIVISKTN 2387
QY 1819 DGNQPEVVOGRNGIDISSAGSKHSVAIGFOAK--AGEAAVAIGRQTOA-GNQSIAIGDNA 1875
Db 2388 TADKENVT---FGLADNINVKDSVVVPGKANGKFGEGAWVNAEDGANGKDGISI---V 2441
QY 1876 QATGQOSTAI-----GTGNV-----VAGKHSGA--IGDPSTVKADNYSYVGNNOQFTDAT 1923
Db 2442 GKGDKDAVAISKDGVTGIIGLTPAGADGNANALIGVNDVKGLDNGDKGNKSTRIV 2501
QY 1924 QTDVFGVGNNTIVTBSNSVALGNSAISAGTHAGTQAK-KSDGTAGTITTAGATGTGVKGF 1982
Db 2502 YTKPNEEEBQV-ATWNDGLVFCDKGTGHEKALGTTTVKVGDDKNIEVEAGDTIRV-L 2559
QY 1983 AQOTAVGAVSVCAS-----GAERRIQN-----VAAGEVSATSTDVANGSOLYKATQGI-- 2030
Db 2560 KDNIDVKGINVTENITVKEGAKINNGNVIDGVADEGVNATSKQAVNGSQLHKVQQQVNN 2619
QY 2031 -ANATNELDHRITHQENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGOGAVAVGL 2089
Db 2620 QATAINKLDHINKVDKDLRAGIAGATAVAFQRPNEAGKSVISLVGVSYSRESAIAVGY 2679
QY 2090 SKLSDNGQVFPKINGSADTQGHV--GAAVG 2117
Db 2680 ARNSDNNKISIKLGGGMSRGDVNFGSIG 2709
RESULT 8
Q8XDGA PRELIMINARY; PRT; 1588 AA.
AC Q8XDGA;
DT 01-WAR-2002 (TReMBLrel. 20, Created)
DT 01-WAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative adhesin (Hypothetical protein).
GN Z5029 OR ECS4480 OR H161.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 0509952;
RA Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,
RA Shinagawa H.;
RT "O157 specific gene similar to H. influenzae adhesin gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE005587; AAG58749.1; -;
DR EMBL; AP002566; BAB37903.1; -;
DR EMBL; AB036416; BAB7814.1; -;

DR PIR; A86036; A86036.
DR PIR; H91188; H91188.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 13.
DR Pfam; PF05662; HIM; 12.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1588 AA; 160150 MW; B2BA4B06BFF28DEC CRC64;

Query Match 9.2%; Score 986; DB 16; Length 1588;
Best Local Similarity 22.9%; Pred. No. 1.2e-23;
Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;

QY 1 MNHLYKVFENKATGTFMAVEYAKSHSTGGSCATGQGVSRITLSPARIALAVLVIGAT 60
DB 1 MNKIFKVPMPATGNTVTTSETAKSRGKSGRSK-----LLSALVAGGM 45

QY 61 LNSAYAGIGISEADGGK---GANARGDKSIAIGDIAQALGSGSIAIGDNKIVHNSNN 117
DB 46 L--SSFGALANAGNDNGQGVYDYGSGSAGDGWALGKGA-----NTEMTSGSS 93

QY 118 ANTG--AKAGNSIAIGGVNLAGSHASIAIGSDLLYKKETVQOIGSELPPIIRGQKALN 175
DB 94 TAVGYDAIARGQYSSAIGSKTHAIGGSMAGF-----VSAISE----- 131

QY 176 DIYQLADTNLQYRTHAQHASTAVGAMSVYAKHFSNAGFTRATAGTYSGLVLTATA 235
DB 132 -----GDRSIALGASSYSLGQYSMALGRYSKALGKUSIAMGDSKKA 172

QY 236 KAASSIAGVNAQAIGFAATGAGSTQVNLNRGIALGFGSOVLQKQNDVNAANVRAYAPD 295
DB 173 EGNALALGNATKATEIMSLALGDUTANASKAYSMLG--ASSVASENAI--AIGARTEAE 230

QY 296 DNPIDNRYKATFNG-----ATDVFSIGNSGNDSIRRKIIINVAGSADTDVAVVA 347
DB 231 NATAIGNNAKAKGTSMAMFGSLADKVNTIALGNSQALADNAIAIGQG-----NKA 283

QY 348 QLEAVPLANRQITFFKGDGNNRVEKGLKTLTITGCAQTSALTDHNIIVVQNGDGLKVQ 407
DB 284 DGVDAALGN-----GSQS-----RGLN--TIALGTASNATGDKSLAGSNS-- 324

QY 408 LAETLSLKMVTENLTANEKVTVGKTRLTATDKIGFTNDMNMGIDESPILDKDTG--IHA 465
DB 325 -ANGINSVALGADSIADLDNTVSGNSL---KEKIVNVKNGAIKSDSY--DAINGSQLYA 379

QY 466 GQKIKITLTA--VVDADAITYGQLKKVQTAESALQTFVKKVQKNGDAN-----DSKI 519
DB 380 ISDSVAKRLGGGAADVDDGT-----TAPTYNLKNGSKNNVGAALAVLDENT 427

QY 520 I---TVGKKNKPDGTQVNTLKLKGENGVDVTTETNGTTFGLNQNNGLTVGNSTLNDG 575
DB 428 LQWDQTKGYSAAHGTSSPTASV-----ITVDAGTIS----- 460

QY 576 LSVKNTNSNQIIOVGADGITFTDISNKPAGIENITRIIRDGIGFANNTGSLDANKPRL 635
DB 461 ASSKDAVNGSLKATNDVE-----ANTANIATNTSNIATNTANIATNTNI 507

QY 636 TPTGINAGKELTNVQSAINPATNGGQLDFPNRLSTANTEKSGSAATIKOLYNLSQVPLT 695
DB 508 T-----NLTDVSGDLQ--ADALLMNETKAFSA----- 534

QY 696 FAGDTGPNVTKLGLKVGKGTADDLTKNNIGVVDSTNSLTIVKLAKTLSDLDVAN 755
DB 535 -----HGQDTISK---ITNVKDA-----DLT-----ADSTD--AVNGSOLKTTNDVATN 574

QY 756 TKTLTASDKVTVDGNNNTAKLQNGDLTFPSKQNTGATPATNSKTIIGVDGLKFTDNNGI--A 813
DB 575 TTNI-----ANNTSNIATNTNINSLTETVT-----NLGEDALKWDKNGVFTA 618

QY 814 LDGTTYITKQVGAQKSDLK--SKPYLDKDLKVGVEITTINGINAGKAITGLSNTL 872

DB 619 AHGTE--TTSKITNVK--DGLTGTGSDAVNGSQLKTTNDVAVNTTNIATNT--TNISLNT 674
QY 873 TDAIT-----ATTGHTVQLGIVDSTDKTRAASIGDVLNAG--FNLKNG 914
DB 675 ETVNLGEDALKWDKONGVFTAAHGNNTASKITNILDGTVTATSSDAINGSQLYDLSSN- 733
QY 915 DAKDFVSTYDVPFINGNATTAKVYDVGKASKVAYDVNDGTTIHLTGAGDNKRNQIGVKT 974
DB 734 ----IATY-----FGNGAS-----VNTDGVFTGPTYKIGETNYNV-- 765
QY 975 TLTITKTDKCD--KAIN--FSVNSGDDKALINAKDIADNLNTLAGEIRNTKGTADTALQT 1030
DB 766 -----GDALAINSSFSLSIG--ALL-----WDATAGKFSAKHGT----- 799
QY 1031 FOVKVKKENG-----DDNDADTITVGDKAKTNOVNTLKLKGN-----GLDIQTNKD 1078
DB 800 -----NGDASVITDVGERT-----SDSSDAVNGSQLHGVSYYVVDALGGAEVNAD 847
QY 1079 GTVTFGINTOSGLKAGNNTTLNNG--LSIKNTAGNEQIOVGADGVKFAKNNGVGAGI 1136
DB 848 GTITAPTPTTIA-----NADYDNVGDALNAIDTTLDDALLWDAD----- 885
QY 1137 DGTTRITRDEIGFAGTNGSLDKSPHLKSDGINAGGKKITNIOSGEIAQNSNDVATGCKI 1196
DB 886 -----AGENGAF--SAAH-----GDKTASVITNVANGAISAASSDAINGSOL 926
QY 1197 YDLKTELENKISSTAKTAQN---SLHEFVSVADEGNNFTVSNPYSSYDTKTSVITFAG 1253
DB 927 YTNKYIADALGDGAEVNADGTITAPTYYTIANAEYN-----VGDAILDALD 972
QY 1254 ENGITTKNVGVVRVIGIDQTKLTPKLTGVNNNGKGIIVDSONGQWTTIGLSNTLANVT 1313
DB 973 DNALLW-----DETA-----NGGAGAYNASHDGKASI-----ITNVA 1004
QY 1314 NDKGSVRTTEQKLIKDEDKTRAASIVDVLSAGENLQNGEAVDFVSTYTVNPAFCNAT 1373
DB 1005 N--GSI-----SEDS-----DAVNGSQLNAT 1024
QY 1374 TAKVTYDDTSTKSKVVYDV--NVDDTTIE-----VKDKLGVKTTTLTSTGTANKF 1423
DB 1025 NMNI-----PQNTQIINQLAGNTDATYIOENGAGINVYRTDDGLAFNDASAGVGATAI 1079
QY 1424 ALSNQATGDALV-----KASDIVAHLNLTSGDIQT---AKGASQANSAGVVDADGNKVI 1475
DB 1080 GYNSVAKGDSVAIGQGSYSDVDTGIALGSSSVSSRVIAGKSRDTS-----ITENGVIIG 1134
QY 1476 YDSTDNKKYQAKNDGTVDKTKVEAKDKLVAQAQTPDGTLAQMNKVSINKQVQNDANKQ 1535
DB 1135 YDTTIDGELLGALSIG-----DGGKYRQ 1156
QY 1536 GINEDNAFVKGLEKAAASDNKTNAAVTVGDL---NAVAQTPLTF--AGDTGTTAKKLGE 1589
DB 1157 IIN-----VADGSEAHDAVTRQLQNAIGAVALTPPKYFPHANSTEEBSLAVGT 1204
QY 1590 TLTIKGQQTNTKLTNNIGVAVGT-----DGFVVKLAKDLTNLNSVAGGTKIDDKG 1642
DB 1205 DSLAMGAKTIVN--GDKGIGIGYGAYVDANALGAIAGSNAQVIHVSIAIGNSTTTTRG 1262
QY 1643 VSF-----VDSGQAKANTPVLISANGLDLGGKVISNVGKTKDTPDAANVQQLNEVERNLL 1696
DB 1263 AQNTYATNMDAPQNSVGEFSVGSAD---GQRQITNVAAGSADTDVAVNGQL----- 1311
QY 1697 GLGNAGNDNADGNQVNIADIKDPNSGSSSNRTVIKAGTVLGGKNNNDTEKLATGGVQVG 1756
DB 1312 -----KYTDAQVSNQTSITN-----LDNRVTNLDSRVT--NIENG 1345
QY 1757 VDKDGNANGDLSNVVKTKQDGSKKALLATYNAGQNTYLTNNPAAIDRINEGIRFFH 1816
DB 1346 I-----GDIV-----TTGSTKYFKN----- 1361
QY 1817 VNDGNQBPVQGRNGIDSSAGKHSVAIGFQAKADGEAAVAIGRQTQAGNQSIAIGNAQ 1876
DB 1362 -----TDGVDASAQGGKDSVAI-----GSGSIAAADN-- 1387

```

QY 1877 ATGDOSIAIGTGNVAGHSGAIGDPSTVKADNSYSGVNNQFTDATQTDVFGVGNITV 1936
Db 1388 -----SVALGTGSV-----ATEENTISVGSSTNQ 1411
QY 1937 TBSNSVALGNSAISAGTHAGTQAK-KSDGTAGTATTTAGATGTGKGFAGTAVGAVSVGA 1995
Db 1412 RRTTVAAGKNA-----TDAVNAQLKSEAGGVRYDTKADGSID--YSNITLG-----GG 1460
QY 1996 SGAERRIQNVAAGEVSATSDAVNGSOLYKATQIGIANATN-----ELDHRIHQENKANAG 2051
Db 1461 NGGITRISNVASG--VNNNDVVNYAQLKQSVQETKQYTDQRVEMDNKLSKTESKLSGG 1517
QY 2052 ISSAMAMAPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGQWFKINGSADTQGH 2111
Db 1518 IASAMAMTGLPQAYIPGASMASIGGTYNGESVALGCVMSVANGRWVYKLGSTNSQGE 1577
QY 2112 VGAAGVAGPHF 2122
Db 1578 YSAALGAGIQW 1588

RESULT 9
Q8GM77 PRELIMINARY; PRT; 1004 AA.
AC Q8GM77;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Adhesin.
GN HIA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=montypable strain 1862A;
RX MEDLINE=22300350; PubMed=12410830;
RA Laarmann S., Cutter D., Juehne T., Barenkamp S.J., St Geme J.W.;
RT "The Haemophilus influenzae Hia autotransporter harbours two adhesive
RT pockets that reside in the passenger domain and recognize the same
RT host cell receptor."
RL Mol. Microbiol. 46:731-743(2002).
DR EMBL; AY078086; AAL79952.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 2.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 1004 AA; 104286 MW; 7806E7E2D4E3D017 CRC64;

Query Match
Best Local Similarity 29.2%; Pred. No. 1.5e-22;
Matches 335; Conservative 142; Mismatches 337; Indels 332; Gaps 47;

QY 1124 PAKYNGVVGAGIDGTTTRITRDEIGPAGTNG-----SLDKSPHLKSDGINAGKK 1173
Db 44 FATVQASV--AATSGTT-----GTNGLHTYGDSFNFNNSATDLNRH-VEDAYK 90
QY 1174 KITNI-QSGEIAQN-----SNDVATGKIKYDLKTELENKISSTAKTAQNSLHEFSVADEQ 1228
Db 91 GLLNLNEKADKSNFLVANDTTATVGNLRLKGLWLSKKNKTRNEKSQQVKHAEVLFEK 150
QY 1229 NNFTSNPYSSYDTSKTD---VITPAGENGITTKVNGVVRVIGIDQTKLTPPKLTG 1285
Db 151 DGVTV-----TSKSENGKHTVTTFLANDLVK-----NATVSDKLSLG- 188
QY 1286 NNGKGVIVDSQ-----NGQNTIIGLNTLANVNTNDKGSVTTTQGLIKIDEDKT 1334
Db 189 ANGKVKDITSADNGLKFAKPGTNGQNGVHLNGIASTLDDPRVGGKTAHLTKETISDTERN 248

```

```

QY 1335 RAASIVDLVSAGFNLO-----GNGEAVDFVSTYDTNFAENATTAKVTYDDTSTSKV 1388
Db 249 RAASVGDVLNAAGNIRGAKTIGTVDNDFVSTYDYVEFASGANANVSTTDDNKITT-- 306
QY 1389 VYDVNVDDTTTIEVK-----DKKLG-----VKTTTILSTG 1417
Db 307 ---VRVDVTGLPVQVVTEDSKTVKVGNEYEYEAQKDSADMDKKVENGKLAKTVKVLVSA 363
QY 1418 TGANKFALSQATGALVKASDIVA--HLNTLSGDIQTAKGASQANSAGYVDADGNKVI 1475
Db 364 NGTNPVKISNADG--TEDTDAVSFKQLKALQ-DKQVTLASNAYANGG-SDADGKVT 418
QY 1476 -----YDSTDNKY--YQAKND-----GTVDKTKVAKDKLVAQATPDGTLA 1515
Db 419 KTLNGLNPKFKSTDGELNLIKAEENDTVTFPKKSV-QVGGDKKASISKGANTTEGLVE 477
QY 1516 QMNVKSVINKEQVNDANKKQGINEDNAFVKLEKASDNKTKNAAVTVVGDLANAVAQTPLT 1575
Db 478 ASELVDSLK-----LGWKVGVDKDS-----GELDGTFKETLV 511
QY 1576 FAGDTGTTAKKLGETLTIKGGQTDNKLTDNIGVVAGTDTGTVKLAKDLTNLSV---- 1631
Db 512 KSGDKVTL--KAGDNLKVK--QEGTN-----FTYALRDELTVGKSVFEKD 552
QY 1632 -----NAGTKLIDDKGVSPVDSG-----QAKANTPVLISANGLDLGGKVISNVKGTGKT 1681
Db 553 TANGANGASTKITKDLGITTPANGANGAATDADKIKVASDGI.SAGNKAVKVVSLKKF 612
QY 1682 DAANYQQLNEVRNLLGLGNAGNDAGNOVNADIKKDPNSGSSNRTVIKAGTVLGGKG 1741
Db 613 GDANF-----DPLTSSADNLT-----KQ 630
QY 1742 NNDTEKLATGGVQGVGDK-----DGNAN--GDLSNV-WVKTKQDGSKALLATYNAAGQ 1792
Db 631 NDDAYKGLTNLDEKGTDKQTPTVAONTAATVADGLGLGWV--ISADKTTGELNKEVNA--- 686
QY 1793 TNYLTNNPAEADIRINEQIRFFHVNDGQEPVVOGRNGIDSS---ASGKHSVAIGFOAK 1849
Db 687 -----QVRNANEVFKSG-NGIHVSGKTVNGRREIT--FEA 720
QY 1850 ADGEAAVAIGRQTOAGNQSIAGIDNAQATGDSIAIGTGNVAGHSGAIGDPSTV--KA 1907
Db 721 KDENA-----IAFGYGSKALRDNTVAITGTVNVVNAEKSGAFGDPNYIEDKA 766
QY 1908 DNSYSVGNNOFTDATQTDVFGVGN-----ITV----- 1936
Db 767 GGSYAFGNDNRIT-SKNTFVLGNGVNAKYKANGVDVTETVTVKDKDGKETTVPKALGA 825
QY 1937 TESNSVALGNSAISAGTHAGTQAKKSDGTAGTATTTAGATGTGKGFAGTAVGAVSVGAS 1996
Db 826 TVNSVYLGNK---STATKDKGNLKSDDGTAGNTTATGTTGTVNGPAGATAHGAIVSVGAS 882
QY 1997 GAERRIQNVAAGEVSATSDAVNGSOLYKATQIGIANATNELDHRHQENKANAGISSAM 2056
Db 883 GEERRIQNVAAGEISATSDAINGSOLYKAVAGVTN---LAGQNVNKKRADAGTASAL 938
QY 2057 AMASHPQAYIPGRSMVTGGIATHNGOGAVAVGLSKLSDNGQWFKINGSADTQGHVGA 2116
Db 939 AASQLPQASMPGKSMVSIAGSSYQGNGLAIGVSRISDNGKVIIRLSGTTNSQSGKTGVAA 998
QY 2117 GAGPHF 2122
Db 999 GVGYQW 1004

RESULT 10
Q8GM78 PRELIMINARY; PRT; 1002 AA.
AC Q8GM78;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Adhesin.

```


QY 1480 DNKYQAKNDGTVDKTEVAKDKLVAQAQTPDGTTLAQMNVKSVINKEQVNDANKKOGINE 1539
 Db 911 ----- 910
 QY 1540 DNFAVKGLEKAAADNKTNAATVVDGLNAVAQTPLTFAGDTGTAKKLGETLTIKGGQTD 1599
 Db 911 -----VTVG-----SGVTLSGMLVI----- 926
 QY 1600 TNKLTNNIGVAGTDTFTVKLAKDLTNLSNVNAGTKIDDKGVSVDSGQAKANPVL 1659
 Db 927 -----TDG-----PSV 932
 QY 1660 SANGLDLGGKVISNVGKGTDDAANVQQLNEVRNLLGLGNAGNDNADQNVNIADIKD 1719
 Db 933 TSSGINAGSQKTNVAAAGTADTDVNLSQLN----- 963
 QY 1720 PNSGSSNRVTIKAGTVLGGKGNNDTEKLT--GGVQGVVDKDGANGDLSNVVWKTQKD 1777
 Db 964 -----TAMASSGAKSVMHYSTVDGGTQ-----CGNYNGD----- 992
 QY 1778 GSKKALLATYNAGQNTYLTNPAPALDRINEGIRFFHVDNQNQPEPVVQGRNGIDSSAS 1837
 Db 993 -----GAT 995
 QY 1838 GKHSVAIGFOAKADGEAAVAIGRQTOA--GNQSIAGDUNAQAATGDSIAIGTG----- 1888
 Db 996 GTGSTAVGVTLASAGATAIGSAAASGKSTAGRNVAASADGSVALGDGAKDGARGA 1055
 QY 1889 NVVAGKHSG-----AIGDPSTVKADNSYSVGNNOFTDAT---QTDVFG-----V 1930
 Db 1056 ESYTGKYSQVQNTVGTVSVDGASKGETRTVSNVADAKEATDAVNLRLQDQVAQDANRV 1115
 QY 1931 GNNI-----TVTESNVALGSNSA--ISAGTHA-----GTQAKSDGTA--GTTTTAGA 1975
 Db 1116 DNKIESLSEGTQFVKVNSV---NSATPIAAGVDATAIGVGATAGSADSIAMGNKASASA 1172
 QY 1976 TGTVGFGAQTAVG-----AVSVGASGERRIONVAAGVSVATSDAVNGSOLYKA 2026
 Db 1173 DNVA-----AIGHSVADRANTVSVGSAGSERQVTNVAAG---TADTDVNVSQL--- 1219
 QY 2027 TQGIANATNELDRIHQNEKANAGISSAMAMSPQAYIPGRSMVTGGIATHNGQGAVA 2086
 Db 1220 NQGLITAKQYTDGVVGLRRDRTDGGVAAAIATANLPQAYIPGRGTSVGVSSVSGQSAIA 1279
 QY 2087 VGLSKLSDQGVFKINGSADTQGHVGAAGVAGFHF 2122
 Db 1280 VGVSSVSGRWFVKFSGSANTRSQVGIGAGVGYQW 1315

RESULT 13

Q98HJ2
 ID Q98HJ2 PRELIMINARY; PRT; 1953 AA.
 AC Q98HJ2;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein ml12848.
 GN ML12848.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RA MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti ";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AF003000; BAB49874.1; --.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR008640; Hep Hag.
 DR InterPro; IPR008635; HIM.
 DR InterPro; IPR000911; Ribosomal_L11.
 DR InterPro; IPR005594; Yada.
 DR Pfam; PF05658; Hep Hag; 30.
 DR Pfam; PF05662; HIM; 8.
 DR Pfam; PF03895; Yada; 1.
 DR PROSITE; PS00359; RIBOSOMAL_L11; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1953 AA; 184557 MW; 892BFA8687B35E2 CRC64;

Query Match 7.8%; Score 831; DB 16; Length 1953;
 Best Local Similarity 22.5%; Pred. No. 1.3e-18;
 Matches 540; Conservative 251; Mismatches 828; Indels 780; Gaps 107;
 QY 17 MAVEA-----KSHSTGGSCATGQV-----GSVRTLSFARIAALAVLVIGAT 60
 Db 39 MAVLRMAHRLVGMRRKSGSLGMLGIRVLAALGAIALGAIATPALAQYAA-----GGGT 93
 QY 61 LNSAYAGIGISEADGG-----KGGANARGDKSIAIGDIAO-----ALGSQSIAGD 107
 Db 94 ANSAGSVAVGPSATTNGLRGIAGVSGATQAGIDISIAQGTSAAGSQAQVSAQVSIASQL 153
 QY 108 NKIVHNSNNANIGAKASGNESIAIGDVLASGHASIAIGSDLLYLKKTETVQ-QISELLP 166
 Db 154 NSIYLGSRTAAGTGATA--QSAIGIGTDTVASQVDIAVGRSSVASAQVSVAGLSAKAT 211
 QY 167 IIRGOKALNDIYQLADTN--LQYRTHAQGHASTAVGAMSYAKGHFSNAGFTRATAEGTY 225
 Db 212 GTGGAWALQGTGTIANSINVALGVQASATGGCANALGTFSVASGGNSTAVGTSSSAAGL 271
 QY 226 SLAVGLTATAKAASSIAGVSNQAQIGFAATAVGGSTOVNLRGIALGFQSGVQLKQNDVN 285
 Db 272 SPAGGWSAVASGDSVALGKSANAAGLNSALGSSNTASADPALA----- 317
 QY 286 AANVRAYAPDDNQPIDNRYKATFKNGATDVFISIGNSGNDSIRRKTIINVAGSADTDVAVN 345
 Db 318 -----GNOAVSSGIGSVAAGSG--SQATG 339
 QY 346 VAQLKEAVRLANRQITFKGDDSNRVERKGLKTLTITGGAQTSALTDRHNVGVVQGDGLK 405
 Db 340 VS-----ATALGNNA-----TAANATALGLGAT--AGGVDCGA-----IGKAN----- 378
 QY 406 VOLAETLTSLKMTVTENLTANEKVTVGKRLTTD-----KIGFTNDMG-----IDESKPYLD 458
 Db 379 ---ASAQDAIAMGTSAKASSAQAIAGVINAATGKAVSIGSGNTAYGDGAVSIGDPSYA 435
 QY 459 KDTGTHAGQKITKLTAGVWDDDAATYQQLKKVNTQTAESALQTFVKKVKDKNGNDANDSK 518
 Db 436 SGTGAFTGANNIANSDDGTATATA-----NOAAGA----- 466
 QY 519 IITVGKNNKP-----DGTQVNTLKLKGE-----NGVDVTTETNGTFTFLGNQNNGLT 565
 Db 467 -VAIGNNKAIGQGSVALGNGSTAGAGLAGNVALNGA--TAAASGSDVALGSGSVTAVA 524
 QY 566 VG--NSTLNNGLSVKNTNSNKQIQGVADGI--TFTDISNKGPGAGIENNTTRITRDGIGF 621
 Db 525 VGTPTNAVINGTITAFQGTPTTPTSVSIGAPGAERTLTNVA-----AG-----RIS----- 568
 QY 622 ANNTGSLDANKPRLTPTGINAGKELTNVQSAINPATNGQGLDFMRLSTANTEKSGSAA 681
 Db 569 GSSTDAVNGSOLFATNQAVDAIGTTLNNIN-----GGGKYPH-----ANSTLADSSA 617
 QY 682 TIKDLYNLSQVPLTFAGDTGNVTKLGEILKVKGGKTTADDLTQKNITGV--VADSTONS 739
 Db 618 TGTDSVAI-----GPTSTATASSAI--AGSNANASGANASAICTSSVASALDAT 665

```
QY 740 LTVKLAKTLDLDAVNTKTLTASDKVTVDSNNNTAKLQNGDLTFPSKO--NTGATPATNSKT 798
Db 666 AMGFSLKASQSFSTAVGANANATLSTALQNALASGVQATLKGQANAGA-----SEA 720
QY 799 IGVNGLKFTDNNGLDGTYYITKDKYGFAGQDGLDKSPYLDKDKLKVGEVEITNGI 858
Db 721 LALGANSTAGNAGDVALGSGSVTAVAG-----TPNAVINGTYYAFQGI 764
QY 859 NAGKAITGLSNTLTATNATTHGVTOLGIVDSTDKTRAASI---GVLNAGFNLKNGD 915
Db 765 NPASTVSIAGPAGERTLTNLAAGRISG-----SSTDVANGSQLPATNOQVDAIGTVNN-- 818
QY 916 AKDFVSTYDVTDFINGNATTAKVTYDGKASKVAYDVNVDTGTHLTGADGNKQIG-VKT 974
Db 819 ----ISTGGGIKYFHANSTLA---DSSA-----TGTD--SVAIGPTST 852
QY 975 TLTUTKDAKDKAIFSVNGDDKALINAKDIADNLN--TLAGEIRNTKGTADTALQTFQV 1033
Db 853 ATASSAIAAGSNA-----NASCANASAITGSSVASALDATAMGFLSKASQGFSTA----- 902
QY 1034 KVKENGDDNDADTTIVGKDAKTNQVNTLKLKKNGLDITQNKDGTVTGINTQSLKA 1093
Db 903 -----VGANANATLSTAI--GQNAL-----ASGVQATLKGQANASA 939
QY 1094 GNNNTLNNGLSIKNTAGNE--QIQVGADGVKPAKVNNGVVGAGIDGTTRITRDEIGFAGT 1152
Db 940 SDALALGAN-----STAGNAGDVALGSGSVTAVAG--GTPNAVINGTIT-----YAFQGI 986
QY 1153 NGSLDKSKPLSKDGINAGG--KKITNIQSGEIAQNSNDVATGGIKYIDLKTELE----- 1204
Db 987 NPA-----STVSIAGPAERTLTNLAAGRISGSSSTDVANGSQLPATNOQVDAIGTV 1038
QY 1205 NKIS-----STAKTQNSLHEESVADQGNFT-- 1232
Db 1039 NNISTGGGIKYFHANSTLADSSATGSDSVAIGPTSTA--TASSAIAAGSNANASANASAI 1097
QY 1233 -VSNPYSSYPT-----SKTS-DVITPAGENG-----ITTKVNGVVRVIGID-----QT 1273
Db 1098 GTSVSASALDATAMGFLSKASQGFSTAVGANANATLSTALQNALASGVQATLKGQA 1157
QY 1274 KGLTTPKLTGVNNGKIVIDSQNGQNTIT-----GLSNTLANVT-----N 1314
Db 1158 NAGASDALALGANSTAGNAGDVALGSGSVTAVAGTPTNAVINGTYYAFQGINPASTVISG 1217
QY 1315 DKGSVRTEQ---GKLIKEDKTPAASIVDLSAGFNQLQNGEAVDFVSTYDVTNFDGN 1371
Db 1218 APGAERTLTNLAGRI-----SGSSTDVAV--GSOLFATNOQVDAIGT--TVNNLGGG 1266
QY 1372 ATT-AKVTYDDTSKTSKVVDVNVDDTTIEYKDKKLGKVTTLTSTGTGANKFALSNOAT 1430
Db 1267 VTNLGNVNNIAGDTSYATDANGIGIRY-ARTNEAGLAQTDSPAQGLGSTAVGYQASAT 1325
QY 1431 GD---ALVKASDIVAHLNTLSGDIQTAKG--ASQANSSA-GYVDA-DGNKVIYDSTNKY 1483
Db 1326 GVSGLALGRGT-----LASIDGVSALGSGSVADRAVPATGQIAAGPANFIQNTYDTKL 1380
QY 1484 YQAKNDGTVTKTEKAVDKLVAQAQTPDGTGLAQMNVKSVINKEOVNDANKKQGINEDNAF 1543
Db 1381 LGAVSVGTATSYRQI-----TNVADGTQAO-----DAVTVRO----- 1412
QY 1544 VKGLEKAASDNKTK--NAAVTGVGLNNAVAQTPLTTFAGDTGTAKKLGETLTIKGGQTDTN 1601
Db 1413 LOGVIASVSATSTKYFHFANSTAGDSLAV-----GAESVAVGPTTVWNG----- 1455
QY 1502 KLTDNNIGVAVG-----TDGPF--TVKLAKDL-----T 1626
Db 1456 ---DNGGIGINGALVDOIPAEGTAIGQNAHVMLADGLALGTNSLASIGQSVAGAGAQT 1512
QY 1627 NINSVNAGGKIDDKGVSFVDSGQAKANT--PVLISANGDLIGG----KVISNVGKGTKOT 1681
Db 1513 FINSVALGAQSVTVVG---AQAGYTAFTAPQTSVGEVSIIGGAGAEKRLTNVAAGSANT 1569
```

```
QY 1682 DAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKDPDNGSGSSNRTVIKAGTVLQKG 1741
Db 1570 DAVNVSQLK-----GVSNVANL-----FGG-- 1590
QY 1742 NNDEKLTATGGVQGVYDQKGNANGDLISNWWKTKQDGSKKALLATYAAAGQTNVLT---- 1797
Db 1591 -----STTVNPDGSI TG-----PTYTIQG--NNYSTVYDG 1618
QY 1798 ----NNPAEAI DRINEQIRFFHVNDGNQBPVVQGRNGIDSSASGKHSVAIGFOAKADGE 1853
Db 1619 FTAVNNALTNIS--NGGGIKYFHANS-----TLADSTAGTDSVAL-----GP 1659
QY 1854 AAVAIGRQTOAGNQSGIAIGNDAQATQDSIAITGTVNVVAGKHGAIGDPSTVKADNSYSV 1913
Db 1660 ASVA-----SGTNSLAAGNGSTATQGAVALGQ-----AKANN----- 1693
QY 1914 GNNNQFTDATQTDVFGVGNNTITVTESSVALGSNSAISAGTHAGTQAKKSDGTAGTTTAA 1973
Db 1694 -----ANDVALGSGSV-----TQ-----TAVGTSST 1714
QY 1974 GATGTVKGFAGQTAVGAVSVGASGAERRIQNVAAAGEVSATSTDAVNGSOLY----- 2024
Db 1715 VINGKTYAFAGTPTGTIVSVGDAGAERTITNVAGRVNAGSTDAINGSOLYATNTAVEDL 1774
QY 2025 ----KATQ-----GIA-----NAT---NE 2036
Db 1775 KSLGLSLTONAVTYDTNPDGSKKNSITLQGGDVNAPVVISNVGPGVAGTDAVNVNOMNR 1834
QY 2037 LDHRIHQN-----ENKANAGISSAMAMASMPQ 2063
Db 1835 VDIATNTSNTSYDKVAATTLQOANNVYTDKLSQNSLDLGIRDEARQAAAGLAASLRY 1894
QY 2064 AYTPGRSMVTGGIATHTNGQGAVALGSLKSDNGQWVFKINGSADTQGHVGAAGVAGPHF 2122
Db 1895 DDRPKLSVAAAGGGFWRDSALAFAGYGTSEDGEIRGVNSGTA--AGGHVG--VGAGISF 1950

RESULT 14
Q8PCQ5 PRELIMINARY; PRT; 2351 AA.
AC Q8PCQ5; 2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XCC0658
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C. de Souza R.F.,
RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DB EMBL; AB012164; AM39974.1; -.
```



```
Db 1689 GVNA--SALGQSAALADNALAIGNSRADAIGASVVGVDASATGINSTGVGRQVNVIGE 1746
Qy 1682 DAANVQOLNEVR-----NLLGIG-NAGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTV 1736
Db 1747 NAVSVGYNSFVRESAVNGVALGANAGATGADSVAL-----GSGSRT-YEANTV 1793
Qy 1737 LGKGKN-----NDTEKLATGGVQGVVDKGNANGDLSNVWVKTKDKGSKKAL 1783
Db 1794 SVSGNGRGGPATRIIVNSDGE-VATDAVNKGQDLADADVQTTSGMVQTTGEGVARAT 1852
Qy 1784 --LATYNAAGTNYLTNNPAEALDRINEQGRFPHVNDGNQEPVV-QQRNGIDSSAGKH 1840
Db 1853 GDRATAAGAT-----ASGARSVAVAAGSTASATGASAMGVDSASGVN 1897
Qy 1841 SVAIFGQAKADGEAAVIGRQT---QAGNQSIALGDNAQATGDSOIALGTG-----NVV 1891
Db 1898 STAMGRQTSIGENGVALGYNFSVFRQSGANAVAGANAGASGADSVAGSGSRTYEANVV 1957
Qy 1892 A-----GKHSIGAIGDPST-----1904
Db 1958 SVSGSNGRGGPATRIIVNVGAGATADASTDAINGGQFFQSLSNTASFLGGGAIGAQGVF 2017
Qy 1905 -----1904
Db 2018 VAPTYVIQASYNVNGAALTALDSKVTELDARGGTATATARTVSLRTAAVPAVASATV 2077
Qy 1905 -----VKADNSVSGVNNNQFTDATOTDFG-----1929
Db 2078 ASSDVTSSASVQGTPTAAVVGVTATPAATSTAVGNAAVANHI TGTATGGSAIYAGPNDTA 2137
Qy 1930 VGNNTVTESSVALGNSAISLA-GTHA-----GTQAKKSDGTA-GTTTATAGATGVKGF 1982
Db 2138 IGSNARVNDGSTAVGANTQIAAVATNAVAMEGAQVSAASGTAIGQARASAOGAVALG 2197
Qy 1983 AQGTA--VGAVSVCASGARRIONVAAGEVSATGTDVANGSOLYKATQGTIANATNELDHR 2040
Db 2198 QGSVADRANTVSVSGVSGEROVANVAAG---TRATDAVNKGQL---DSGVAAANSYTDSSR 2251
Qy 2041 I-----HONENKANAG-ISSAMAWAMPQAVIIPGRSMVTGCIATHN 2080
Db 2252 YSAMADSFETYQGDIEDRLRQRNRLDRQAGAMGAMLNMSASVAGIASQNRIGAGVGFQN 2311
Qy 2081 GQGAVALGSLKSLDNGQVWFKINGSADTQGHVGAAGVAGF 2120
Db 2312 GESALSVMGYQRAISPRATV--TIGGALSGDDSSIGVGVAGF 2349
```

RESULT 15

```
Q8GM75 PRELIMINARY; PRT; 1210 AA.
AC Q8GM75;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Adhesin.
GN HIA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
CX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=nonotypable strain 3179B;
RX MEDLINE=22300350; PubMed=12410830;
RA Laermann S., Cutler D., Juehne T., Barenkamp S.J., St Gene J.W.;
RT "The Haemophilus influenzae Hia autotransporter harbours two adhesive
RT pockets that reside in the passenger domain and recognize the same
RT host cell receptor.";
RL Mol. Microbiol. 46:731-743(2002).
DR EMBL; AY078089; AAL79954.1; --
DR InterPro; IPR008635; HIM.
DR InterPro; IPR001680; WD40.
```

```
DR InterPro; IPR005594; YadaA.
DR Pfam; PF05662; HIM; 4.
DR Pfam; PF03895; Yada; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
SQ SEQUENCE 1210 AA; 125073 MW; 6A26A860A947FEA9 CRC64;

Query Match 7.1%; Score 763; DB 2; Length 1210;
Best Local Similarity 24.6%; Pred. No. 1.1e-16;
Matches 356; Conservative 170; Mismatches 451; Indels 470; Gaps 65;

Qy 855 TNGINAGGKA--ITGLSNLTLDAT-----NATTGHTVLTGLVDSIDTKTRAA 898
Db 55 TNGINAYGSTNPFADALNNSATDLNRVNDPAYSGLLNLTNEKGTANKQLLVADST---AA 110
Qy 899 SIGDVLNAGNLKNGDAKD---FVSTVDTVDFT-NGNATTAKTVYDGK-----ASKVA 948
Db 111 TVGDLRLKLVVSTKSTKEESNQVKADEVLTGSGAATVSSKSENGKHTITVSVTKGE 170
Qy 949 YD-VNVVDTTTHLTGADGNKNQIGVKT---TTLTKTDAKDKAINFVSNSGDDKALINA 1003
Db 171 FNTVKTTDAATAGAAGA-NTNERGKVTVSGVTDLTKTAE-----DKKQVATV 216
Qy 1004 KDIADNLNTLAGEIRNTKGTADTALQTF-QVKKVKENGDDDDNDADTITVGDKATNQVNT 1062
Db 217 VDVAKAINDAA-----TFVKVENSNEELDDNQAD---NADQALKAGDT 257
Qy 1063 LKLG-KGNGLDIOQNKDG-TWTFGINTQSGLKAGNNTLLNNGLSIKNTAGNEOIQVGAD 1120
Db 258 LTFKAGKN---LKAKRDKKNIITAL-----ANDLNVTATVSDKLSLGAN 299
Qy 1121 GVKPAKVNVGVGAGIDGTTTTRITRDEIGFA---GTNSLDKSKPHLSKDGINAGGKITN 1177
Db 300 G--NKNV-----VTSITT---DGLNFAKETAVNGD-----TN 325
Qy 1178 IQSGEIAQNSDAVTCGKIYDLKTELENKISSAKTAQNSLHSPSVADQGNFTVSNPY 1237
Db 326 IQLSGIGSTLTDITG-----ITKTATN-----348
Qy 1238 SSYDTSKTSDVITFAGENGITTKVNGKVRVGDIDQTKGLTTPKLTGVNNGNGKVIDSQN 1297
Db 349 -----GVAVENHN 356
Qy 1298 QONTITGLSNTLANVNDKGSVRTTEGKI IKBEDKTRAASIVDVLISAGFNLQNGEA-- 1355
Db 357 -----LAASVKDVLISAGNIRGAATAGG 379
Qy 1356 ----VDFVSTYDTVNFA-DGNATTAKTVYDDTSKTSKVVDVNVDDTTIEVKDKKLGVKT 1410
Db 380 VDNVDVSTYDTVDFTDDGETTITVTYQKANGKGAEV--KIGAKTSVIEKDGKLVTKG 437
Qy 1411 TLTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTSLGDIQTAKGASQANSAGYVD-- 1468
Db 438 ANKEAGGTNTTP-ATEDMDGKGLVTAETVINAVNAGWRIKT-----TTANGQAGKETV 492
Qy 1469 ADGNKVIYD-----STDNK--YQAKNDGTV-DKTK-EVAKDKLVQAQATPDGTLAQ 1516
Db 493 TSGTNTVTFDNGNATTAIVSKDKGNIIVKYDVNVVGDGLKVDATKIVA-----DITALT 547
Qy 1517 MNVKSVLNKEQVNDANKKQGINEDNAPVKGLEKASDNKTKAAVTVGDLNVAQAQPLTF 1576
Db 548 VN-----SGNEANKPKG-----KVAEITNEGDKKKLVNABGLVNALNSLSWT--AT 591
Qy 1577 AGDTGTTAKKLGETLTIKGGQTDNTNKLTDNNIIGVAVGTGDTFTVKLAKDLTNLSV---- 1631
Db 592 AGKDGTVNAGSADQEVKAGDKVTFKAGD-NLEVKQDGNFTVSLNSTLGLTSITLTKT 650
Qy 1632 -NAGGTKIDDKGVSFVDSGQA-----KANTPVLISANGLDGGKVISNVGKGTQDADAANV 1686
Db 651 ANGATTKITKDGTLTTPANGCATGTNNANIISVTTSGISAGNAKAITNVASS----- 701
Qy 1687 QQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKGNNDTE 1746
Db 702 -----LNAYGDT-NTNFDATANSATDLTROFDANGAYNGLL-----NLNE 740
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:32:39 ; Search time 39.4567 Seconds
(without alignments)
2800.358 Million cell updates/sec

Title: US-09-813-214A-9
Perfect score: 10708
Sequence: 1 MNHIVKIFNKATGTFMAVA.....NGSADTQGHVGAAGAGFHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	630.5	5.9	2249	1	OMPA_RICRI	P15921 rickettsia
2	550.5	5.1	2021	1	OMPA_RICCN	Q52657 rickettsia
3	545	5.1	2660	1	YEEJ_ECO57	O8x8v7 escherichia
4	544	5.1	2003	1	YDBA_ECOLI	P33666 escherichia
5	508	4.7	2358	1	YEEJ_ECOLI	P76347 escherichia
6	494	4.6	1856	1	OMPB_RICJA	O06653 r outer mem
7	493.5	4.6	1953	1	BIGA_SALTY	P25927 salmonella
8	480.5	4.5	3590	1	FHAB_BORPE	P12255 bordetella
9	480	4.5	1643	1	OMPB_RICPR	Q53020 r outer mem
10	474.5	4.4	1577	1	HLVA_PROMI	P16466 proteus mir
11	466.5	4.4	2432	1	Y43R_IRV6	P18305 chilo iride
12	456.5	4.3	1655	1	OMPB_RICCN	Q9Kka3 r outer mem
13	444.5	4.2	1902	1	P3P_LACLC	P15292 lactococcus
14	437.5	4.1	1902	1	P1P_LACLC	P16271 lactococcus
15	437.5	4.1	1902	1	P2P_LACLC	P15293 lactococcus
16	435.5	4.1	1654	1	OMPB_RICRI	Q53047 r outer mem
17	433.5	4.0	1902	1	P2P_LACPA	Q02470 lactobacill
18	430	4.0	1645	1	OMPB_RICTY	P96389 r outer mem
19	425	4.0	1608	1	HLVA_SERMA	P15320 serratia ma
20	403.5	3.8	1569	1	YPJA_ECOLI	P52143 escherichia
21	398	3.7	1848	1	CBPA_CLOCL	P38058 clostridium
22	380.5	3.6	2334	1	WAPA_BACSU	Q07833 bacillus su
23	366	3.4	1325	1	YDEK_ECOLI	P32051 escherichia
24	360	3.4	1300	1	120K_RICRI	P14914 rickettsia
25	354	3.3	1176	1	SLAP_BACSH	P38537 bacillus sp
26	353	3.3	1829	1	FRPC_NEIMC	P55127 neisseria m
27	351	3.3	1839	1	ALE3_AZOVI	Q44496 azotobacter
28	349.5	3.3	1829	1	FRPC_NEIMB	Q9Jyv5 neisseria m
29	342	3.2	1723	1	PM20_CHLPN	Q92812 chlamydia p
30	339	3.2	1025	1	SLAP_CAUCR	P35828 caulobacter
31	338	3.2	1286	1	AIDA_ECOLI	Q03155 escherichia
32	329.5	3.1	1694	1	IGA0_HAEIN	P44969 haemophilus
33	325.5	3.0	1567	1	ICEN_XANCT	P18127 xanthomonas

ALIGNMENTS

RESULT 1

ID	OMPA_RICRI	STANDARD	PRT	2249 AA.
AC	P15921			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Outer membrane protein A precursor (190 kDa antigen) (Cell surface antigen) (rOmpA) (rOmpA).			
DE	OMP.A.			
GN	Rickettsia rickettsii.			
OS	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;			
OC	Rickettsiaceae; Rickettsia			
OX	NCBI_TaxID=783;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=R;			
RX	MEDLINE=90354033; PubMed=2117568;			
RA	Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;			
RT	"A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";			
RL	Infect. Immun. 58:2760-2769(1990).			
CC	-!- FUNCTION: ELICITS PROTECTIVE IMMUNITY.			
CC	-!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.			
CC	-!- PTM: Glycosylated (Probable).			
CC	-!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP.B FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M31227; AAA26380.1; -.			
DR	PIR; A41477; A41477.			
DR	InterPro; IPR006315; Autotransport.			
DR	InterPro; IPR005546; Autotransporter.			
DR	Fam; PF03797; Autotransporter; 1.			
DR	TIGRFAM; TIGR01414; autotrans_bar1; 3.			
KW	Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.			
FT	SIGNAL 1 28 POTENTIAL.			
FT	CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.			
FT	DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.			
FT	REPEAT 212 286 A (TYPE I).			
FT	REPEAT 287 358 B (TYPE II).			
FT	REPEAT 359 430 C (TYPE II).			
FT	REPEAT 431 505 D (TYPE I).			
FT	REPEAT 506 577 E (TYPE II).			
FT	REPEAT 578 652 F (TYPE I).			
FT	REPEAT 653 724 G (TYPE II).			
FT	REPEAT 725 799 H (TYPE I).			
FT	REPEAT 800 874 I (TYPE I).			
FT	REPEAT 875 949 J (TYPE I).			

FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match 5.9%; Score 630.5; DB 1; Length 2249;
Best Local Similarity 23.2%; Pred. No. 5.4e-16;
Matches 500; Conservative 212; Mismatches 779; Indels 665; Gaps 109;

QY 180 LADNLOKRYRTHAQ-----HASTAV-----GAMSYAKG-----HFSNAPGTR- 218
D 1 MANISPLKFKAIQGLKAALFTTSTAAILMSSSGALGVATGVIATNNAAFSNNVGN 60
QY 219 -----ATAGT-----YSLAVGLTATAKAASSIAGVNAQAIGFAATAVGST 261
D 61 WNETAAGVANGTPAGGPQNNWAFYGGDYVTADAADRIKAIN-----VAGTT 110
QY 262 QVNLN-----RG-----IALGFGSOVLQKNDVNAANVRAYAPDDNQPIDNRYK 305
D 111 PVGLNITQNTVVGSIITKGNLLPVTNAGKSLTLNGNNAVAANHGFDPADNY----- 163
QY 306 ATFKNKATVPSIGNSGNDSIRKII NVGAGSADTDAAVNAQLEAVRLANROITPKGD 365
D 164 -----TGLNIALGGANAAL-----IIQSAAPSKITLAGNID-----GGGIITVKTD 205
QY 366 DSNRVEKGLKTLTITGGAOTSAITDHNIGVQNGDGLKVQLAE-----TLTSLKMTT-- 420
D 206 AALNGTTGNTNALATVNVGAGATLG-----GAVIKATTTKLTNAASVLTLTNANAVLTGA 261
QY 421 -ENLTANEKVTV-----GKTRLTDTKIGFTNDMGIDESKPYLDKDTGIGHAGGQKITLTA 475
D 262 IDNTTGGDNGVGLNGLALSQVTDIGNTSLATISVG-----AGTATLGGAVIKATT 314
QY 476 GVVDHDAATYQLKKNQOT-----ABSALQTFVKVKKDKNNDANDSKIIIV-----G 523
D 315 TKLTDAASAVKFTNPVNVVTGAIDNTGNANNGIVTFTGNST-VTGNVGNNTALATVNVGAG 373
QY 524 KNNKPDG--TOVNTLKLKENGVDVTEFTGVTFTGL-----NQNGLT--VGNSTLANDG 575
D 374 LLQVQGVVKANTINL--TDNASAVTTPVNVVTGAIDNTGNANNGIVTFTGNSTVIGD- 430
QY 576 LSVKNTNSKQIOVG-----ADGITFTDISNSKPGAGIENTTRI 614
D 431 --IGNTNALATVNVGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGA-IDNTT-- 485
QY 615 TRDIGFANWTGSLDANKPRLTPTGINAGGKELINVQSAINPATNGQGLDFMNLSTANT 674
D 486 GGDNVGVLNGLALS-----QVTGNIGNTSLATISVGAGTATLGGAV-----IKATT 534
QY 675 EKSGSAATIKDLYNLSQVPLTFAGDTGNVTKLGELIKVGGKTTADDLTKN-----IG 730
D 535 KLTDAASAVK--FTNPVVTGAIDNTGNAN--NGIVTFTGNSTVIGDIGNTSLATIS 588
QY 731 VVADSTNSLTSLVLAKTLSLDVAVNTKLTASDKV---TVDS-----GNNATKLO-NGDLTF 783
D 589 VGAGTATLGGAVIKATTTKLTNAASVLTLTNANAVLTGAIDNTTGGDNGVGLNGLALS- 647
QY 784 SKONTGATPATNS-KTIGVDGLKFTDNNGIALDGTYYITKDKVGFAGQDGLSKPKYLD 842
D 648 --QVTGDIIGNTSLATISV-GAGTATLGGAVIKATTTKLTNAVASAVK-----FT 693
QY 843 KDKLVGVEITTINGIN-----AGGKAITG-LSNTLTDTAT-NATTGHVTLQGLIV---DST 892
D 694 NPVVVTGAIDSTGNANNGIVTFTGNSTVIGDIGNTNALATVNVGAGTATLGGAVIKATT 753
QY 893 DKTRAASGTDVNLGNAGFNKNNGDAKDFVSTVDVDFINGNATTAKVTYD--GKASKVAYDV 951
D 754 KLTNAASVLTLTNANAVLTG--AIDNTTGGDNGVGLNGLALSQVTDIGNTSLA-TI 809
QY 952 NVDTGTHLTGADGNKNOIGVKTTLTKTDAKDKKAINFSVNSGDDKALINAKIADNIN 1011
D 810 SVGAGTATLGG-----VIKATTTKIT-----NAASVLTLTNANAV----- 845

QY 1012 TLAGEIRNTKGTADTALQTFQVKVKEKNGD--DNDADATTITVGKDAKT-----NQVNTLK 1064
D 846 -LTGAVDNTTGGDNGVGLNGLALSQVTDIGNTSLATISVGAGTATLGGAVIKATTTK 904
QY 1065 LKGNKGLDIOTNKDGTVTTFGINTOSGLKAGNNT--TLNNGI-----SIKNTAGNEQIOV 1117
D 905 LTNAAASVLTLTNANAVLTGAIDNTTG--GDNVGLNGLALSQVTDIGNTSLATISV 961
QY 1118 GA-----DGVKFAKVNNGVVGAG-IDGTRITRDEIGFAGTN--GS 1155
D 962 GAGTATLGGAVIKATTTKLTDAASAVRFTPNPVVTGAIDNTGNANNGIVTFTGNSTVGN 1021
QY 1156 LDKSKPHLSKDGKINAGGKITTNIQSGEIAQNSNDAVTGKTIYDLKTELENKISSTAKTAQ 1215
D 1022 VGNINA-LATVNVGAG--LLQVQGVVKANTINLTD-----NASAVTFT 1062
QY 1216 NSLHFSVADEQGNNTVSNPYSDYTSKTSDDVITFAGENGIT-----TKNKGKV 1266
D 1063 NPVVVTGAIDNTGN-----ANNGIVTFTGNSTVTVGNVGNNTALATVNVG-- 1106
QY 1267 RVGIDQTKG-----LTTPKLTVG-----NNNGKGIIVDSQNGQNTIT 1303
D 1107 -AGLLQVQGVVKANTINLTDNASAVTFTPNPVVTGAIDNTGNANNGIV--TFTGNSTV 1163
QY 1304 G-LSNTLANVTNDKGSVRTEQGGKIIKDEDKTRAASIVD-----VLSAGFNLOQNGEAVD 1357
D 1164 GDIGNTNALATVNVGAGITLQAGGSL-----AANNIDFGARSTLEFNGPLDGGKAIP 1216
QY 1358 F-----VSTYDTVNFADGNATTAKTYDDTSTKTSKVY 1390
D 1217 YFFKGAINGANNAILNVNTKLLTASHLTIGTVABINIGAGNLTIDASVGD-----VT 1269
QY 1391 DVNVDDTTIEVKDKLGVKTTTLTSTGTGANKPALSNOATGDALVKASDIVA----- 1442
D 1270 ILNAQNTNFRARDSVLVLSNL-----TGVGVNILL-----AADIVAPADEGTV 1314
QY 1443 ----HLNLTSGDIIOTAKGASQANSAGYVDADGNK-----VIYDS--TDNKYYQAKNDGT 1491
D 1315 VFNGGVNGLN-----VGSNVAGTARNIGDGGKNKENTLLIYNAVTTTDDVLEGIQNVL 1368
QY 1492 VDKTKEVAK-----DKLVAQATPDGTL--AQMNVKSVINKEQUNDA 1531
D 1369 INKNADFTSSATFNAGAIQINDATYTTIDANNGLNIPAGNIQFAHADAQLVLQNSSGNDR 1428
QY 1532 NKQGINEDNAFVKGLEKAASDNKTNAAVTVGDLNAV-ACTPLTEAGDTGTT---AKKL 1587
D 1429 TITLGNID-----PDNDEGIVI-----LNSVTAGKKLTIAG--GKTFGGAHL 1471
QY 1588 GETLTIKG-GQTDNTKLTNNIGVAGTGDFTVKLAKDLTLNLSVNAVAGTAKIDDKGVSVF 1646
D 1472 -QTILFKGAGDCST-----AGTTENTTVILDT----- 1499
QY 1647 DSSQQAANTPVL SANGLDLGGKVISNVGKTKTDDAANVQQLNEVRNLLGLGNAG---N 1703
D 1500 --GC-----LELG-----ATTANVLFNDVQLTGTGNIIGGFLD 1531
QY 1704 DNAD-----GNQVNIADIKKDPNSGSSNRTVIKAGTVLGGCKGNNDTEKLA---TGQVQ 1754
D 1532 FNKNGMVTLLNNVNVAGAVQ--NTGGTNNGTLI-----VLGASNLNRVNGIAMLKVGN 1585
QY 1755 VGVDKDGNNAN-GDLSNVVWVKTKQDGSKKALLATYNAAGQTNLYLTNNPABADRINEQIR 1813
D 1586 VTIKAGGKVKIGEIGTGTNT-----LTLPAHENLTG-----SINKTGGQALK 1628
QY 1814 PFHVNDGNQBPVQVRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAGD 1873
D 1629 LNFVNGSGSVGV-----GTAANSVGDIITTAGATSFASBV 1663
QY 1874 NAOAT-----GDQSLA-----IGTGNVNVAGKHSIGAIGDPSTVKADNSYSVGNNOFTDATQTD 1926
D 1664 NAKGTATLGGTTSFANTFTNTGATLAK-----GSITSFAKNVTAT 1704

QY 1927 VFGVGNITVTESNVALGNSAISAGTHAGTQAKKSDGTAGTTTGA-----TGTVKG 1981
 Db 1705 SP-VANSATINFSNLAFTSN--ITGG-----GTTTLGANGVYTTGT-GS 1746
 QY 1982 FAGQTAVGVSGA--SGAERRIQNVAAAGVSTATSDAVNGSQYKATQGIANATN 2035
 Db 1747 FDTTLTLNTFDCAAKSGGNILKSGSLDLSGVSTLAL-----VVTATN 1791

RESULT 2
 OMPA_RICCN STANDARD; PRT; 2021 AA.
 ID AC Q52657; P95591; P95593; P95594; Q52667; Q52668; Q52669;
 AC Q52670; Q52674;
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
 antigen) (rOmpA) (rOmp A).
 GN OMPA OR RC1273.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=94171067; PubMed=8125327;
 RA Croquet-Valdes P.A., Weiss K., Walker D.H.;
 RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
 conorii (Malish 7 strain).";
 RL Gene 140:115-119(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098(2001).
 RN [3]
 RP SEQUENCE OF 8-204 FROM N.A.
 RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
 RX MEDLINE=97015921; PubMed=882558;
 RA Roux V., Fournier P.E., Raoult D.;
 RT "Differentiation of spotted fever group rickettsiae by sequencing and
 analysis of restriction fragment length polymorphism of PCR-amplified
 DNA of the gene encoding the protein rOmpA.";
 RL J. Clin. Microbiol. 34:2058-2065(1996).
 RN [4]
 RP SEQUENCE OF 953-2012 FROM N.A.
 RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
 RA Raoult D., Fournier P.E., Roux V.;
 RT "Phylogenetic analysis of spotted fever group rickettsiae by study
 of the outer surface protein rOmpA.";
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 layer with hexagonal symmetry.
 CC -!- PTM: Glycosylated (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.
 CC
 DR EMBL; U01028; AAA17405.1; -.
 DR EMBL; AB008674; AAL03811.1; -.

DR EMBL; U43794; AAB49549.1; -
 DR EMBL; U43798; AAB49550.1; -
 DR EMBL; U43806; AAB49551.1; -
 DR EMBL; U45244; AAB49566.1; -
 DR EMBL; U46918; AAB48663.1; -
 DR EMBL; U83440; AAC35176.1; -
 DR EMBL; U83443; AAC35179.1; -
 DR EMBL; U83448; AAC35184.1; -
 DR EMBL; U83453; AAC35189.1; -
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR005546; Autotransporter.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRFAMs; TIGR01414; autotrans_barl; 1.
 KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
 KW Complete proteome.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 2021 OUTER MEMBRANE PROTEIN A.
 FT DOMAIN 238 946 THR-RICH.
 FT DOMAIN 1424 1528 THR-RICH.
 FT VARIANT 60 60 N -> NN (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 76 76 R -> H (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 86 137 MISSING (IN STRAIN M1).
 FT VARIANT 126 133 MISSING (IN STRAIN MOROCCAN).
 FT VARIANT 953 954 VT -> II (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 1245 1245 D -> A (IN STRAINS INDIAN TICK TYPHUS, M1
 AND MOROCCAN).
 FT VARIANT 1308 1308 N -> H (IN STRAIN MOROCCAN).
 FT VARIANT 1877 1877 M -> I (IN STRAIN INDIAN TICK TYPHUS).
 FT CONFLICT 10 10 Q -> K (IN REF. 1).
 FT CONFLICT 92 92 I -> V (IN REF. 1).
 FT CONFLICT 126 126 V -> I (IN REF. 1).
 FT CONFLICT 137 137 T -> N (IN REF. 1).
 FT CONFLICT 157 157 G -> D (IN REF. 1).
 FT CONFLICT 368 369 IS -> VN (IN REF. 1).
 FT CONFLICT 374 388 KATLGAIKATTTK -> LLQVGGVVKANTIN (IN
 REF. 1).
 FT CONFLICT 640 640 N -> D (IN REF. 1).
 FT CONFLICT 669 669 V -> I (IN REF. 1).
 FT CONFLICT 793 793 N -> D (IN REF. 1).
 FT CONFLICT 803 804 VN -> IS (IN REF. 1).
 FT CONFLICT 809 823 LLRVQGVVKSNTIN -> KATLGAIKATTTK (IN
 REF. 1).
 FT CONFLICT 898 898 D -> V (IN REF. 1).
 FT CONFLICT 908 908 P -> N (IN REF. 1).
 FT CONFLICT 985 985 N -> K (IN REF. 1).
 FT CONFLICT 1009 1009 L -> S (IN REF. 1).
 FT CONFLICT 1013 1013 Y -> S (IN REF. 1).
 FT CONFLICT 1182 1182 K -> Q (IN REF. 1).
 FT CONFLICT 1314 1314 N -> Y (IN REF. 4).
 FT CONFLICT 1451 1451 H -> N (IN REF. 1).
 FT CONFLICT 1624 1624 G -> D (IN REF. 1).
 FT CONFLICT 1628 1628 E -> G (IN REF. 1).
 FT CONFLICT 1872 1872 A -> V (IN REF. 1).
 FT CONFLICT 1875 1875 T -> P (IN REF. 1).
 FT CONFLICT 1878 1879 MS -> LP (IN REF. 1).
 FT CONFLICT 1936 1936 E -> A (IN REF. 1).
 FT CONFLICT 1965 1970 MTAPLP -> ITTPPLS (IN REF. 1).
 FT CONFLICT 1997 1997 G -> R (IN REF. 1).
 SQ SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;
 Query Match 5.1%; Score 550.5; DB 1; Length 2021;
 Best Local Similarity 21.5%; Pred. No. 4.1e-13;
 Matches 511; Conservative 243; Mismatches 811; Indels 811; Gaps 116;
 QY 14 GTFMAVEYAKSHSTGGSCATGQGVSVTLTAFARIALAVLVIGA-----T 60
 Db 88 GDYTTADVADHIIITAINVADTTPIG---LNAQNTVGVSVTGGNLLPVITAGKSLT 143
 QY 61 LINGSAGYIGISEADGGKGGANARGDKSTAIGDIA-----QALGQSSTIAIGNKIVHNSN 115
 Db 144 LNG-----NNADAANHGFAPADNYTGLGNIALGGANAALIIQSAAPAKITLAGNIN 195
 QY 116 NNANIGAKASGNESIAIGG-DVLASGHASIAIGGDDLYLKKEVTQVQISSELLPIRQKAL 174

Db 196 GGGIITVKTDAINGTIGNTNALATVNVGAGIATLEGAIKATTKLTNAASVL----- 249
Qy 175 NDIYQLADTNLQKRRTHAQGHASTAVGAMSAYAKHESNAPGTR-----ATAGTYSLAVG 230
Db 250 -----TLTNVAVLTGAI DNTTGV DNVGVNLNGALSQVTG 285
Qy 231 LTATAKAASSIAGVSNAAQI GFAATAVGGSTOVNLRGIALGFGSQVLQK-----DNDVNAA 287
Db 286 NIGTNALATISVGAGKATLGGA--VIKATTKLT DNASAVTFNPNVVTGAIDNTGNA-- 342
Qy 288 NVRAYAPDDNQPIDNRKYATPKNGATDVFSIGNSGNDSIRKLIINVAGSAD----- 340
Db 343 -----NNGIVTFTGDSVTGNTGNATNA-----LATISVGAGKATLGGAIIK 383
Qy 341 -----TDVNVAVOLKEAVRLA-----NQIITFKGDDSN-----RVEKG 374
Db 384 ATTKLT DNASAVTFNPNVVTGAI DNTGNANNIGIVTFTGDSVTGIGNTNALATISVG 443
Qy 375 LGKTLTITGG---AQTSAITDHNIGVQVQNGDGLKVQLAETLTSIKMVT-----TENLTAN 426
Db 444 AGKA--TLGGALIKATTKLT DNASAV-----TFNPNVVTGAIDNTGNAN--N 488
Qy 427 EKVTVGKTRLTDDKIGFTNDMNGIDESKPYLDKDTGTHAG-----GOKITKLTAGVDDDA 482
Db 489 GIVTFTGDSVTGNTGNALATI-----SVGAGKATLGGAIIKATTKLT DN 537
Qy 483 ATYQOLKKVQOT-----AESALQFT--VKKYDKNNGDANDSKILTVGKNNKP--DGTQ 532
Db 538 SAVTFNPNVVTGAI DNTGNANNIGIVTFTGDSVTGIGNTNALATISVGAGKATLGAI 597
Qy 533 VNTLKLKXGNGVDVTTETNGVTTFGL-----NONNGLT--VGNSTLNDGSLSVKNTNSNK 585
Db 598 IKATTKLT DNASAVTFNPNVVTGAI DNTGNANNIGIVTFTGDSVTG--NIGNTNALA 654
Qy 586 QIQVGA-----DG-----TFPTDISNSKPGAGIENTRITRDRDIGGFANN 624
Db 655 TVNVGAGIATLEGAVIKATTKLTNAASVLTLTNVNAVLTGA--IDNTTGV--DNVGVNL 711
Qy 625 TGSJLD-----ANKPRLTPTG INAGKELTNVQSAI--NPATNGGOL 663
Db 712 NGALSQVTGNTGNTNALATISVGAGKATLGGAIIKATTKLT DNASAVTFNPNVVTGAI 771
Qy 664 DFMNRLSTANPEKSGSAATIKDLYNLSQVPLTFAGDTGPNVTKLGE-----I 711
Db 772 D-----NTGNA--NNGIA-----TFTGDS--TVTGNIGNTNALATSVNVGAGL 809
Qy 712 LKXVGGKTADDLTKNNGI VVADSDNSLTVKLAKTLSLDDAVNTKLTASDKVTVDSG-- 770
Db 810 LRVOG-----GVVKSNTIN-----LTD-----NASAVTFNPNVVTGAI 843
Qy 771 NNTAKLQNGDLTFGKQON--TGATPATNS--KTIGYDGLKFTDNNGIALDGTYYITKDKVGF 827
Db 844 DNTGNANNIGIVTFTGDSVTGNTGNALATISVGAGKATLGGAIIKATTKLT DN 900
Qy 828 AKODGSLDKSPYLDKOKLKVGEVEITTINGINAGGKAITGLSNTLT DATNATGHVTQLG 887
Db 901 ----SAVTFNPNV-----VVTGAIDNTGNANNIGIV-----TFTGDSVTG 936
Qy 888 IVDSTDKTRAASIGDVNLNAGNLKNG--DAK--DFVSTYDVTDFINGNATTAKVTVDGK 943
Db 937 NIGTNALATVNVG-----AGVTLQAGGSLDANNIDF--GARSTLEF--NG-----PLDGG 983
Qy 944 ASKAYADVNDGTTIHLTG--ADGNKQIGVKTTLTKTDKAGDKATNFVSNSGD----- 996
Db 984 QNAIPY-----YFGKATANGNALLNVNTKLLTAYHLITGTVAEINIGAGNLFAID 1034
Qy 997 ----DKALINAKDIADNLNTLAGBIRNTKGTADTALQTFQVKVKENGDDN-----DADTI 1049
Db 1035 ASAGDVTLTNAQDI--HPRAL-----DSALVLSNLTGUVN-----NILLAADLV 1077
Qy 1050 TVGKDAKTQNVNTKLKXKNGLIDOTNKGDTVTFGINTQSGLKAGN--NTTLNNGLSIK 1107

Db 1078 APGVDECTVVD-----GGVNLNIGSNVAGAA-----RNTGVDGNGKFNNTLII NAVTIT 1128
Qy 1108 NTAGNEQIQVAGDGVEKFAKVNNGVVGAGIDGTTITRDEIGFAGTNGSLDKSKPHLSKDG 1167
Db 1129 D-----DYNLEGIQNVLINN---NADFTSSAFNAGTTQINDATYIIDANNGLN--- 1175
Qy 1168 INAGGKITNIQSGEIAONS--ND-AVTGGKIYDLKTELENKISSTAKTQNSLHEFSVA 1224
Db 1176 IPAGNIFAHADAQILLQNSSGNDRTITLGANIDPDNDDEGIVILNSVTA----- 1225
Qy 1225 DEQGNFTVSNPYSSYDTSKTSVDITPAGENGITTKYKNKGVVRVIGIDOTKGLTTPKLTVG 1284
Db 1226 ---GKLTIIAGGKTFGGAHLQDIV--FKGE-----GDFGTAGTTF----- 1261
Qy 1285 NNNKGIVIDSQNGQNTITG---LSNTLVNTNDKGSVRTTEQOKIIKDEKTRAAISVD 1341
Db 1262 --NTTNIVLD-----ITQLELGATTANVVLFKDAVOLQTGT----- 1296
Qy 1342 VLSAGFNQNGEAVDPVSTYDVTNVPADGNATTAKVYDVTDSKTSKVYVDVNDVTIEV 1401
Db 1297 ----NIGG-----FLDFNAKNGTTLNN-----NVNVAGT----- 1322
Qy 1402 KDKKLVKTTTLTSTGT---GANKFALSNOATGDALVKASDIIVAHLNTLSGDIQTAKGA 1457
Db 1323 ----VKNTGGTNGTLLIVLGASNL---NRVNGIAMLKVG-----AGNVTIAGK-- 1363
Qy 1458 SOANSSAGYVDADGNKVYDSTDNKYQAOKNDGTVDKTEKVAKDKLVAQAOTPDGTIAQM 1517
Db 1364 --GNVKIGELOGTGTNTL---TLPAHF--KLITGSINKT-----GGQALKL 1401
Qy 1518 NVKSVINKEOVNDANKKQINEDNAFVKGLEKAASDNKTKNAAVTVGDLNNAVQTPITFA 1577
Db 1402 N-----FMNG-----GSVGVVGTAAANSV 1420
Qy 1578 GDTGTT--AKKLGELTITKG---QDTDNKLTOMNIGVWAGTDTGTVKLAKDLT--- 1626
Db 1421 GDITTAGTAFASSVNAKGTATLGGTTSFAHTFTNTGAVTLAKGSIITSFAKNVTATSFVA 1480
Qy 1627 NLNSVNAG-----GTKIDDDKGV-----SFVDS-----SQAKANTPVL 1659
Db 1481 NSATINFGSLAFNSNITGSGTTLTLCANQVYTGTSFTDTLTLNTTDFDGAAGSGGNIL 1540
Qy 1660 SANG--LDLGGKVISNVGKGTKTDDAANVQOLNEVRNLLGNGNAGNDAGNQVNIADIK 1717
Db 1541 IKSGSTLDSLQ--VSNLALVVVAT-----NFDMN----- 1567
Qy 1718 KDPNSGSSNRRTVIKAGTVLGGKG-----NND-----TEKLATG 1751
Db 1568 ---NISPTKVTVISAEATAGLKPENVKITINNDNRVDFDFTDASTLTLPADIAAG 1624
Qy 1752 GVQGVDKDGNANGDLSNV-----WVKTKDGSKK-----ALLATYNAAGOT 1793
Db 1625 ----VIDEDFAPGCPLANIPNAANI KKSLELMEDAPNGSDARQAFNFGMLTLPLEADAT 1680
Qy 1794 NYLTNN---PAAIDRINEQOI-----RFFHNDGNOEPVQVQNGIGDSSAS 1837
Db 1681 THLMQDVVKPSDITAAVNNQVVASNISNITARNMDKVQAGNKGPSVGDEMD----- 1736
Qy 1838 GKHSVAIGFOAKADGEAAVAGIROTAQGNOSIAGDNAQATG-----DQSLAIGTG 1888
Db 1737 -----AKFGAWISPPFVGNATQKNCNIS--GYKSDTTGGTIGTGFDFGVFSDDLVLGLA 1785
Qy 1889 NVVAGK-----HSGAIGDPSTVKAD-----NSYSYGNNNQFTDA----- 1922
Db 1786 YTRADTDIKLKNNTGDKNKVESNIYSILGLYSVPYENLFVEALASYSDNKIRKSRVRI 1845
Qy 1923 ---TOTDVFVGNNITVTESNVAL-----GNSAISAGTHAGTQAKKSDGTAGT 1969
Db 1846 ATTLETYGYQTANGKYKSESYTQLMAGYTYMMSENINLTPLAGLRYSTIKDKSYKETGT 1905
Qy 1970 TTTAGATGTVKGFAGQTAGVAVSVGASGABERRIONVAAGEVSATSDAVNGSOLYKATQ-- 2028
Db 1906 TY---QNLTVKKNYNTFDGLLGAKVS-----SNINVNEIVLT-----PELYAMVDY 1949

QY 2029 GIANATNELDHRTHONENKANAGISSAMAMPOA 2064
 Db 1950 AFQKYSADARLQ-----GWTAPLPTNSFKQS 1977

RESULT 3
 YEEU ECO57
 ID YEEU ECO57 STANDARD; PRT; 2660 AA.
 AC Q8X8V7; Q8X2B9; Q8X2C0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeeJ.
 GN 23135 OR ECS2775/ECS2776.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Ferna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=0157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22 (2001).
 CC -!- SIMILARITY: Contains 16 Big-1 domains.
 CC -!- SIMILARITY: Belongs to the intimin/invasin family.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 1315.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB005423; AAC57041.1; -
 CC EMBL; AF002559; BAB36198.1; ALT_FRAME.
 CC EMBL; AF002559; BAB36199.1; ALT_FRAME.
 CC InterPro; IPR003344; Big_1.
 CC InterPro; IPR003535; Intimin.
 CC InterPro; IPR008964; Invasin_intimin.
 CC Pfam; PF02369; Big_1; 16.
 CC PRINTS; PR01369; INTIMIN.
 CC SMART; SM00634; BID_1; 16.
 CC SMART; SM00089; PKD; 8.
 CC Hypothetical protein; Repeat; Complete proteome.
 CC DOMAIN 738 834
 FT DOMAIN 738 834
 FT DOMAIN 840 929
 FT DOMAIN 931 1033
 FT DOMAIN 1042 1132
 FT DOMAIN 1134 1236
 FT DOMAIN 1245 1335

FT DOMAIN 1337 1439 BIG-1 7.
 FT DOMAIN 1448 1539 BIG-1 8.
 FT DOMAIN 1548 1652 BIG-1 9.
 FT DOMAIN 1653 1750 BIG-1 10.
 FT DOMAIN 1751 1855 BIG-1 11.
 FT DOMAIN 1856 1957 BIG-1 12.
 FT DOMAIN 1963 2056 BIG-1 13.
 FT DOMAIN 2065 2156 BIG-1 14.
 FT DOMAIN 2157 2252 BIG-1 15.
 FT DOMAIN 2254 2355 BIG-1 16.
 SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 5.1%; Score 545; DB 1; Length 2660;
 Best Local Similarity 20.9%; Pred. No. 9.1e-13;
 Matches 489; Conservative 306; Mismatches 888; Indels 652; Gaps 114;

QY 178 YQLADTN---LQYRR-----THAQGHASTAVGAM--SYA-KGHFSNAGFTRA 219
 Db 412 FDLVDRNNNIVLEVRKKELVRLTLTDPVTGKSGEVKSLVSSLTQKVALKGY--NVEATAL 469
 QY 220 TAEGTYSLAVG-----LTATAKAASSIAVGSNAQAIGFAATAVGSGTQVNNLRGIALGFGS 275
 Db 470 EAAGGKVVTTGDKDILVTLTPAYRFTSTPTDNTWPTEVTAEDVKGNFS--NREQSMVVVQAP 528
 QY 276 QVLQKDNVNAANVRAYAPDDNQPIDNRYKATFKNGATDVFSIGNSGNDSIRKLIINVG 335
 Db 529 TLSQKSSVSLSS-----QTLSDSHSTATLTFIAHDAAGNPVIGLIVSTRH 575
 QY 336 AGSADTDVNVVAQLKEAVRLANRQITFKGDDSNRRVEKGLGKTLTITGGAQTSALTDHNI 395
 Db 576 EGVQD---ITLSDWKD-----NGDGSVTQI-----LTTGAMSGTLT--L 609
 QY 396 GVVQNGDGLKVQLAETLTSKMTTENTLANEKVTVGKTR-LTTDKIGFTNDMNGIDSK 454
 Db 610 MPQLNG-----VDAAKAPAVVNIISVSSRTHSSIKIDKDYLSGNPIEVTVELRD-ENDK 664
 QY 455 PYLDKDTGIHAGGQKTKLTAGVVDADAATYQGLKKVNOTABESALQTFVTKVKVDKNGNDA 514
 Db 665 PVKEQKQQLNT-AVSDINVKPGVTTDWKETADGVYKATYATYTKGSLGTLAKLLMQWNED 723
 QY 515 NPSKIITVGKNNKPDGTQVNTLKLKGENGVDTTETNGVTVF-----GLNQNNGLTGVNS 569
 Db 724 LHTAGFIIDAN--POSAKIATLS--ASNGVLANENAAVTVSVNVADGSGNPINDHVTFA 780
 QY 570 TLNNDGLSVNTSNKQIQVGADGTFITDLSNKPAGIENTRITRDG-----IGFA 622
 Db 781 VLSGSATSPNNQNTAK---TDVNGLATFDLKSQKE---DNTVEVTLENGVKQTLIVSFV 834
 QY 623 --NNTGSLDANKPR-----LTPTGINAGGKELTNVQSAINPATNGGQDPMNR 668
 Db 835 GDSSTAQVDLQKSNVADGNDSTMTATVRDAKNLNDVKVTNVNSAAKLS-----890
 QY 669 LSTANTKSGSAATIKDLNLSQVPLTFAGDTGPNVTKLGEILKVKGGKTKTADTLTKN 728
 Db 891 QTEVNSHDGIAATLTSKN-----GDY--TVT-----ASVSSGSGQANQ 928
 QY 729 IGWVADSTNSLTVLAKLTSLDLDAVNT-----KTLTASDKVTVDSGNNTAKLONGLTFS 784
 Db 929 VIFIGDQSTAALTSLVPS--GDITVTNTAPLHMTATLQDK-----NGN---PLDKKEITFS 979
 QY 785 KQNTGATPATNSKTIIGVDGLKFTDNGIA-----LDGTTVITKDKGVKFAQDGSGLDKS 837
 Db 980 VFNDAVASRFSISNS-----GKGMTDSNGTIAASLTGTLAGHMITA-----RLANSNVSDT 1030
 QY 838 KP---YLDKDK-----LKVGEVEITTINGINAGGKAITGLSNTLTDAITNATTHVHTQGIYD 890
 Db 1031 QPMWTFVADKRAVVVLQTSKAEIINGVDE-----TTLTATVKDPPDNV---VKLSV- 1081
 QY 891 STDKTRAAISIGDVLNAGFNKNGDA-----KDFVSTYDTVD--FINGN-----A 933
 Db 1082 --FRTSPADTQLSLNAR--NTNENGAIEAVTLKGTVLGVHTAEAILLNGNRDRTKIVNIAPDA 1138
 QY 934 TTAKYTYDQKASKAVDVNVVDGTTIHLTGADGNKQIGVKTTTLTKTDAKDKAINFSV- 992

CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000237; AAC74483.1; ALT SEQ.
DR EMBL; AE000237; AAC74487.1; ALT SEQ.
DR EMBL; D90778; BAA15009.1; ALT SEQ.
DR EMBL; D90778; BAA18980.1; ALT SEQ.
DR EMBL; D90779; BAA18981.1; ALT SEQ.
DR EMBL; X62680; -; NOT ANNOTATED_CDS.
DR EcoGene; EG11307; ydbA.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12CB53220EE CRC64;

Query Match 5.1%; Score 544; DB 1; Length 2003;
Best Local Similarity 21.3%; Pred. No. 7.1e-13;
Matches 415; Conservative 242; Mismatches 695; Indels 600; Gaps 102;

QY 274 GSQVLQKNDVNAVRAPDNPQIDNRYA-----TFKQGATVDFIHSNGND 325
DB 122 GDEIIPDDPDDPTPPKPSFNDVILDKTEKTLTIRDSVFTYENADGTISLODSNGRK 181
QY 326 SIRRKIINVAGSADTDANVAQLKEAVLANRQITFKGDDSNRVEKGLGKTLTITGCA 385
DB 182 A-----TINLQWIDE--ANNVVALEG-----VSADG 205
QY 386 QTSALTDHNGVQVQDGLKVLQALAEITLSLKMVTENTLANEKVTVGKTRLTDD-KIGFT 444
DB 206 ATKQVNHNGELVITGD-----NATVNN--GKT-ITVDGKDSGT 241
QY 445 NDNGIDESKPYLDKDTGTHAGGQKITKLTAGVDDDAATYQGLKKVNOTAESALGTFTV 504
DB 242 TEING-NNGKVIQDGLDVGSGGHGID-----ITGDSATVDN-----KGTMTV 283
QY 505 KKVQDNGNDANSKIITVGNKKNKPDGTQVNTLKLKGENGVDTTETNGVTTFCLNNGGL 564
DB 284 TPESMGIQDGAIV--NNEGESTIINGGTGTQINGDDATANNNGKITVDGDKDSTGT 340
QY 565 TVGNSTLNNDGLSVKNTSNKQIQVADGITFTDISNSKPGAGIENTTTRITRDGIGFANN 624
DB 341 EI-----NNGNKVIQDG-----DLQVSGGGHGD--ITGDS-ATVDN 375
QY 625 TGSIDANKPRLTPTGINAGKELTNVQSAINPATNGGQLDFMRLSTANTEKSGSAATIK 684
DB 376 KGTMTVTDPE--SIGIQVDQDAVNVNEGESAITNGG-----TGTQINGDDATAN 423
QY 685 DLYNLQVPLTFRAGDTPGNVTKKLGRIKLVKGGKTTADDLTKNIGVWADSDTNSLTVKL 744
DB 424 ---NNGKTTVDGKDSGTETIAGNNGKVIQ-----DGLDVGSGGGHGDITGDSATV-- 471
QY 745 AKTSLDLDVNTKTLTASDKVTVDSGNNTAKLQN-GDLTFSKQNTGA-----TPATNSK 797
DB 472 -----DNKGTMVTVDPESIGIQIDQDAIVNNEGESTIITNGGTGTQINGDDATANNSG 524
QY 798 TTGVDLKFTDNNGIALDGTITYTKDKVFAKQDGLSKSKPYLDKDKLKVEVEITNG 857
DB 525 KTTVDG--KDSGTGTIAG-----NIGVNLIDGSL-----TVTGG 556
QY 858 INAGGKAITGLSLNLTDAFNATTHVTLQIGVDSDTKTRAASIGDVLNA-GFNLKKNGDA 916
DB 557 AH-----GVENIGDNGTVNNKGD I-----VWSDT-----GSGIVLINGEGATVSNITGDV 600
QY 917 KQFVSTYDVTDFINGNATTAKVTYDGKASKAVDVNVVDGTTIHLTGAD--GNKNQIGVKT 974

601 N-----VSNEATGFSITTSNGKVSILAGSMQVGDFS--TGVDLNGNNSV----- 642
QY 975 TTLTKTDAK--GDKAINSVNSGDDKALINAKDIADNLTLAGETINTKGTADTALQTFQ 1032
DB 643 -TLAAKDLKLVGQKATGINV-SGD-----ANTVNITGNVLVDKDKTADNAAEYFF 690
QY 1033 VKVKKEN--GDDNDADATTIVGDKAKTNOVNTLKLKGRNGLDIQTNK-DGTVTFFGINTQS 1089
DB 691 DFSVGINVGSNN-----VTLDGKLVVSDSEVTSR-----QSNLFDGSA-----EKTIS 735
QY 1090 GLKA--GNNTTLNNG--LSIKNTAGNEQIOVGADGVKFAKNNG-----VVGAG----- 1135
DB 736 GLWITGDGNTVNMNGGLEIGEKAL-----ADGSOVTSILRTGYSYTSVIVVSGESS 787
QY 1136 --IDGTTTRTD-EIGFAGTNGSLKSKPHLSKDGINAGGKKITNIQSEIAQNSNDAT 1192
DB 789 VYLDGDTTISGEFFLPGAGV-----IRVQDKALLEIGSG----- 821
QY 1193 GGIYDLKTELENKISSAKTQNSLHEFSVADEQGNFTVSNPYSSYDTSKTSDV--IT 1250
DB 822 -----ATLTMQDIDSFEHHG-----TRTVEIQNLG 846
QY 1251 FA-----GEGITTKVKNKVVVRGIDQTKLTTPKLTVGNNGK-----GIVIDSQNGQNTI- 1302
DB 847 FAFVTGEN--TTGINSGTISLLQNGKDPAPSPIVLLATNGGSATNAGTITGVTEQHSVF 904
QY 1303 ----TGLSNTLANVTNDKGSVETTEGKIIKDEKTRAASIVDVLSAGNLQNGEAVDF 1358
DB 905 NKYSTGTSNFI--FNNDVSSI--TGLVAQSN-----STIINTDSGIIDLYGRG-SVGM 953
QY 1359 VSTYDTVNFADGNATTAKVTYD--DTSKTSKVYVDVVDVDTTIEVKDKLGVKTTTLTST 1416
DB 954 LAIADSTAENQKITLDSMWVDANDTTAMEDIASNAIDFGT-----GVGVGTDYS 1005
QY 1417 GTGANKFALSQATGDALVKASDIVAHLNLTSGDIQTAKGASQANSAGYVDADGNKIY 1476
DB 1006 GAGKNATAI-NQLGG-----VITIYNAGAG--MAAYGAS-----NTVIN 1041
QY 1477 DSTDKYIYQAKNDGIVDKTKEVAKDKLVAQAQTPDGTTLAQMVKSVINKEQVNDANKQG 1536
DB 1042 QGTIN--LEKNGNYDSS--LAANTLVGMVVEHGT-----AINDQGTVIN 1082
QY 1537 INEDNAFVKGLEKASDNKTNAAVTVGDL-----NAVAQPLTFAGDTGT 1583
DB 1083 INVGTG-----QAFYNDGTGTIVNYGTICTFGVCQSGNEYNNTDDTSLIYTG--GDT 1133
QY 1584 AKKLGELTITIKGGTDTNKLTDN--NIGVAGTGTFTVKLAKDLNLSNVNAGG-----T 1636
DB 1134 ITRSGETVTLNKSAAVTDKLAGNVVNSGTLSG-DQITVS-----SGLLENTSGGIINLV 1187
QY 1637 KIDDKGVSVFSDSGQAKANTPVLISANGLDLGGKVISNVCKTKD-----TDAANVQQLN 1690
DB 1188 KLD-----KGAVIKNAGVMT--NNVDVSGGILNNAEMTAQITMNAAGUSSLVNNTG 1237
QY 1691 EYRNLLGLGNAGNDAGNQVNIADIKKDPNSGSSNRTVIKAGTVLGGKNNDTKELAT 1750
DB 1238 TINKI--VQNAVFN-----NSGVTGRM-----SAGGVFNQTD----- 1271
QY 1751 GGVQVGVKDGKNGANDLSNVW-VKTKQDGSKALLATYNAAGTN---YLTNNPBAIDR 1806
DB 1272 GAIMRGAALTGTAVANNEGTVNLGSSSEGNNTGMLEVNNSAFNNRGEFILDNDKNAV-H 1330
QY 1807 INEQGI-----RFFHVNDGN-----QEPVVOGRNGIDSSA 1836
DB 1331 INQSGTLYNTGHNMTSNSHNGAVNMWGNGRF--INDGTIDVSAKSLVSVANNAGDQNA 1388
QY 1837 S--GKHSVAIGFOAKADGEAAVAI-----GRQOAGNQSIAT-IGD-NAQATGDOS 1882
DB 1389 FFWQDQNGVINFD--HDSASAVKVTHTSNFIAQNDGIMNISIGTAVAMEGDKNAQLVNGT 1446
QY 1883 IAGTGNVAVGKHSIGAIG-----DFTVKADNSYSVGNNGQFTDATQTDVFGVGNNI 1934
DB 1447 INLGATGTT--DTGMIGMQLDANATADAVIENNGTINIFANDSFAFSVLGTGVGVVNG 1503

QY 1935 TVTESNSV-----ALGSNSAISAGTHAG-----TQAKKSDGTAGTTT 1972
 Db 1504 TVVIADGVGTGLIKQSDSINVEGMNGNSSEVHYGDTYLPDVPKPNVTSVSGDEA 1563
 QY 1973 AGA-----TGT-VKGFAGOTAVGAVSV-----GASGAERRIQNVAAGEVSA 2012
 Db 1564 GGSNNLNGVVGTVNGVSGAKLVNNSMGMVEINTGFTAGTADTTVSPDNVVEGS-NL 1622
 QY 2013 TSTDVNGSOLYKATQIGIANATNELDRIHON 2044
 Db 1623 TDADAITSTVWVTAKGSTDASGNDVDTMSKN 1654

RESULT 5
 YEEJ_ECOLI
 ID YEEJ_ECOLI STANDARD; PRT; 2358 AA.
 AC P76347; P94750;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeeJ.
 GN YEEJ OR B1978.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12";
 RT Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9097040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
 RA Takada J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map";
 RL DNA Res. 3:379-392 (1996).
 CC -1- SIMILARITY: Contains 13 Big-1 domains.
 CC -1- SIMILARITY: Belongs to the intimin/invasin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000289; AAC75042.1; ALT_INIT.
 CC EMBL; D90837; BAA15800.1;
 CC EMBL; D90836; BAA15799.1; ALT_INIT.
 CC EcoGene; EGI3378; yeeJ.
 CC InterPro; IPR003344; Big_1.
 CC InterPro; IPR003535; Intimin.
 CC InterPro; IPR008964; Invasin_intimin.
 CC InterPro; IPR002482; LysM.
 CC InterPro; IPR000601; PKD.
 CC Pfam; PF02369; Big_1; 13.
 CC PRINTS; PR01369; INTIMIN.
 CC SMART; SMO0634; BID_1; 13.

DR SMART; SMO0257; LysM; 1.
 DR SMART; SMO0089; PKD; 6.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834
 FT DOMAIN 840 931
 FT DOMAIN 932 1033
 FT DOMAIN 1042 1137
 FT DOMAIN 1146 1237
 FT DOMAIN 1246 1350
 FT DOMAIN 1351 1448
 FT DOMAIN 1449 1553
 FT DOMAIN 1554 1655
 FT DOMAIN 1661 1754
 FT DOMAIN 1763 1853
 FT DOMAIN 1855 1950
 FT DOMAIN 1952 2053
 FT CONFLICT 105 105 S -> G (IN REF. 2).
 SQ SEQUENCE 2358 AA; 248599 MW; 232249750BPF631ED CRC64;

Query Match 4.7%; Score 508; DB 1; Length 2358;
 Best Local Similarity 20.6%; Pred. No. 1.8e-11;
 Matches 444; Conservative 285; Mismatches 835; Indels 588; Gaps 101;

QY 178 YQLADTN---LQYRR-----THAQGHASTAVGAM--SYA-KGHFSNAFGTRA 219
 Db 412 YDLVDRNNIVLEVRKKELVRLTLTDPVTCGSEKSLVSLQTKYALKGY--NVEATAL 469
 QY 220 TAGTYSLAVG-----LTATAKAASSIAGSNAQAIGAFAATAVGSGTQVNLRGIALFGS 275
 Db 470 EAAGGKVVTGKDILVTLPAYRFTSTPETDNTWPIEVTAEDEVKGNLS-NREQSMVVVQAP 528
 QY 276 QVLQKQNDVNAANVRAYAPDNDQPIDNRYKATFKNGATDVFSIGNSGNDSIRRKINVG 335
 Db 529 TLQKQSSVSLST-----QTLNADSHSTATLTFIAHDAAGNPVGLVLSRTH 575
 QY 336 AGSADTDVAVNVAQLKEAVRLANRQITPKGDDSNNRVEKGLGKTLITITGGAQTSALTDHNI 395
 Db 576 EGVQD---ITLSDWKD-----NGDGSYQI-----LTTGAMSGTLT---L 609
 QY 396 GVVQNGDGLKVQLAEILTSLKMTTENLTANEKVTVGKTR-LTTDKLGFINDMNGIDESK 454
 Db 610 MPQLNG-----VDAAKAPAVVNIISVSSSRTHSSIKIDKRYLGNPIEVTVELRD-ENDK 664
 QY 455 PYLKDGTGIIHAGQKTKITAGVVDVDDAAVYQOLKKNQTAESALQFTVKKVDKNGDA 514
 Db 665 PVKEQKQQLN-NAVSIDNVKPGVTTDKETADGVYKATYATYKSGSLTAKLLMNNED 723
 QY 515 NDSKIITVGKNNKPDGTQVNTLKLKGENGVVDVTTETNGTVTF-----GLNQNNGLTVGNS 569
 Db 724 LHAGFIIDAN--PQSAKIATLS-ASNNGVLANENAAVTSVNVADEGSNPINDHTVTFA 780
 QY 570 TLNNDGLSVKNTNSKQIQVGADGITFTDISNKGAGAGIENTTITFDG-----TGEA 622
 Db 781 VLSGSATSFNNQNTAK---TDVNGLATFDLKSQKE---DNTVEVTLENGVKQTLIVSVFV 834
 QY 623 --NNTGSLDANKPRLPTPTGINAGKELTNVQSAINPATNGQGLDFMNLSTANTEKGS 680
 Db 835 GDSTAQVDLQKSK---NEVVADGNSVMTATVRAKGNLLNDVMTVFNVSAAEKLQ 891
 QY 681 ATTLDYLNISQVPLTFAGDTGPNVTKLGEILKVKGKTKTADDTLTKNIGVADSTNSL 740
 Db 892 TEVNSHDGIATATLTSKNGDYRVT-----ASVSSGSOANQVNFQDQSTAAL 940
 QY 741 TVKLAKTSLDLDAVNT-----KTLTASDKVTVDSNNNTAKLQNGDLTFSKNTGATPATNS 796
 Db 941 TSLVPS--GDIIVTNTAPQVMTATLQDK-----NGN-----PLKDKETITFENVDA----SK 987
 QY 797 KTIQVGLKFTDNNNGIA---LDGT---TYITKDKVGFPAQDGLSKSKP---YLDKDK-- 845
 Db 988 FSIISNGKGMWDSNGVAIASLTGTLAGTHMIMARLA---NSNVSDAQPMTFVADKDRV 1043
 QY 846 --LKVGEVEITNGINAGGKAITGLSNTLTDATNATTVHTVTLGIQIVDSTDKTAAASIGDV 903


```
Db 1044 VVLQTSKAEIINGVDE-----TTLTATVKDPSPNHPVAGIT-----VNFTMPQDVA----- 1089
QY 904 LNAAGFNKNGDAKDFVSYDYDVFINGNATTAKTATKYTDGKASKVADVNVDGTTIHLTGA 963
Db 1090 --ANFTLENGIA-----ITQANGE---AHVTLKGKKA-----GT--HTV-- 1122
QY 964 DGNKNGIQVKTTLTKTDAGKDAKNFNSVNSGDDKALINAKDIADNLNLTLAGEIRNTKGT 1023
Db 1123 -----TATLGNNTSDSQPVTF-----VAD-----KAS 1145
QY 1024 ADTALQTFQVKVKENGDDNDADTTIVGKDAKTQNVNTLKL-----KGKGLDITQNKDGT 1080
Db 1146 AQVVL---QISDEITGNGVDSATLATVKDQDNEVNNLPVTFSSASSGLTLTPGVSN 1202
QY 1081 VTFGI--NTOSGLKAGNNT---TLNNGLSIKNTAGNEQIOVGADGVKAKVNNGVVAG 1135
Db 1203 NESGIAQATLAGVAFGEKVTASLANNGASDNKTVHFIGDTAAAKTIELAPVDSII--- 1259
QY 1136 IDGTTTRITRDEIGFAGTNGSLOKSPHLSKDGKINAGGKKITNIQSGEIAQNSNDVATGGK 1195
Db 1260 -----AGT-----PQ-----NSSGSVIT-----ATVVDNNGFPVKVYT 1287
QY 1196 IVDLKTLENKISSTAKTAQNSLHFSVADQGN-NFTVSNPYSSYDTSKTSIVITPAGE 1254
Db 1288 V-----NFTSNAATAEMWNGGQAVTNEQGKATVTVNTRSTESGARPDVTEASLE 1338
QY 1255 NGITT-----KYNKGVVRVCIQDTKGI-----TTPKLTVGNNGKGIVIDSQNGQN 1300
Db 1339 NGSSLTSTISNVNADASTAHLTLQALFTVSAGETTSYIEVKDNGNGV-----POQ 1392
QY 1301 TITGLSNTLANVNDKGSVRITTEQGIKIDEDKTRAASIVDLVSAGFNLCQGEAVDFVS 1360
Db 1393 EYTLVSPSEGVTPSPNAIYTT-----NHDGNFYASTATKAGVYQL----- 1434
QY 1361 TYDVTNFAQNAFTAKVYDDTSKTSKVVYDVNVDDTTIEVKDKLGVKTKTTLTSTGTGA 1420
Db 1435 ---TATLENGDSMOQTVPVVPVNAEAITLAASKDPVIADNND-----LTLTATVADT 1485
QY 1421 NKFALSQATGALVKASDIVAHLNLTSGDIQTAKGASQANSAGYVDADGNKVIYDSTD 1480
Db 1486 EGNAIANTEV-----TFTLPEDVK-----ANFTLSDGGKVIYDA-- 1519
QY 1481 NKYYQAKNDGTVDKTKEVA-----KDKLVAQAQ---TPDGTLAQM 1517
Db 1520 -----EGKAVTLKGTKAGATHTVASTMTGKSQBLVNVFNFIADTLTAQVNLNVIEDNFAN- 1574
QY 1518 NVKSVINKEQVNDANKKQGINEDNAFVKGLEKAASNKTKNAAVTVGDNLNAVAQTPLETA 1577
Db 1575 NVGMTRLOATVTDGNGNPLANEAVTFLPADVSASFTLGQGGG-AITDINGKAEVTL--- 1630
QY 1578 GDTGTTAKKLGELTIKG-QQDITNK-----LTDNNIGVVAGTDGFTVKLA 1622
Db 1631 --SGTKSGTYPTVSVNNYGVSDTKQVTLIADAGTAKLASLTSVSVFVSTTEGAT--MT 1686
QY 1623 KDLTNLNSVNAAGTKIDDKGVSPVDSGQAKAN-----TPVLSANGLDLG 1667
Db 1687 ASVTDANGPNVBEIKVNFRTSVLSTSVETDGRGFAELVTSVEVGLKTVSASLADXP 1746
QY 1668 GKVISNVGKGTDKTDAAVQVQNLNEVNNLLGLGNAGNDNADGNQVNIAD---IKKDPN--- 1721
Db 1747 TEVISRLNASADVNSGATITSL-----EIEPGQVMVAQDVAVKAVHNDQF 1791
QY 1722 -----SGSSSNRITVIXAGTVLGGKGNNDTEKLATGGVQVGVDDKXGN--ANGDLS 1768
Db 1792 GNPVAHQPVTFSAEPPSSMIISQNTV-----STNTQGVAB--VTMTPERNGSMYVKASLP 1844
QY 1769 N-VWVTKQXDG-----SKKALLATYNAAGQTNLYLTNNPABAIKINEQGIHF----- 1814
Db 1845 NGASLEKQLEAIDEKLTLTASSPLIGVYAPTGA TLATLTSANGTP-VEQOVINFVSYPE 1903
QY 1815 -----FHVNDGNEPVPVQGRNGIDSSASGKHSHVAIQFQAKADGEAAVAIGRQTOAGN 1866
Db 1904 GATLSGGKVRTNSGGQAPVV-----LTSNKGVTYTVTASF-----HNGVTIQTQT--- 1948
```

```
QY 1867 QSTAGDNAAQATGQDSIAIGTGNVAGKH--SGAIGDPSTVKADNSYSVGNNOFTD--- 1921
Db 1949 -----TVKVTGNSSTA-----HVASFIADPSTIAATN-----TDLSTL 1981
QY 1922 -ATQTDVFGVGN---NITV---TESNSVALGNSAISAGTHAGTQAKSKDGTAGTTTGTAG 1974
Db 1982 KATVED--GSGNLIIEGLTVFVFKSGSATLTSLTAVT-----DQNGIATTSVKGA 2029
QY 1975 ATGTVKGFAGQTAVGAVSVG---ASGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQGIA 2031
Db 2030 MTGSVTVSVAVTTAGGQMTVDITLVAGPABTOSVLKSNRSLKGDYTDSAELRLVL--- 2085
QY 2032 NATNELDHRHONENKANAGISSAMAMAPQAYIPG-----BSMWTG---GIA 2077
Db 2086 -----HDSGNPIKVSSEMEFVGQGTNPYIKISAIDYSLNININGDYKATVTGGGEGIA 2138
QY 2078 T-----HNGQGAVALGSLKLSDN--GQWFKING-----SADTQGHVGA 2114
Db 2139 TLIPLVNGVHQAGLSITTIQFTRAEDKIMSGTVSVNGTDLPTTTTFFPSQGFTGA 2190

RESULT 6
OMP_RICUA STANDARD; PRT; 1656 AA.
ID _OMP_RICUA AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOmpB)
DE (rOmp B) [contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptidel. OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]_TaxID=35790;
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -.
DR InterPro; IPR006315; Autotransport.
DR Pfam; PR03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall. 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1338 32 kDa BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
```



```
Db 1140 L-----PSSGSDTRTSDTGFTAG-----TLANGTETLNGDVDMG 1177
Qy 1275 GLTTPKLTGVNNGKVIDSQONTITGLSNTLANVTNDKGSVRITTEQOKIIKDEBKT 1334
Db 1178 G-----WLYNEAGSLTV---NGVTIVGGANALAN-----1205
Qy 1335 RAASIVDVLSAGNQLQNGEAVDFVSTYDTV-NFADGNATT-----AKVTY-----DDTS 1383
Db 1206 -----YGTILDA-DAISTWHSIFNEADGSIITDILLTNGDVTFYNGDFTG 1249
Qy 1384 KTSKVVD---VNVDDTTFIEVKKL-----GVKTTTLTSTGTGANKFALSNQATGD 1432
Db 1250 SIAGTSYQOEIVNTGDMTVAEDGKSLVSGSFYFNEEDATLTNGSAVE-----GSENTII 1305
Qy 1433 ALVKASDIVAHLNLTSGDIQTAKGASQANSAGYVDADGNKVIYDSTD-----1480
Db 1306 NLTRANDSLTQVN---SGTITATNGYSALTITVNGSNDP---KWIWNTATGVINGINPDAPL 1360
Qy 1481 ---NKYQAKNDGTVDTKTKEVAKDKLVAAQATPDGTTLAQMNVKSVINKEQVN---DANKK 1534
Db 1361 INLCRGYNFGNGIIN---VQGDNAVA---ISGTSY--VINLVNSGTINVTGEQKE 1411
Qy 1535 QGINEDNAF-VKLEKASDNKTNAAVTVGDLNNAVAGTPLTFAG-----1578
Db 1412 DGTNGTGLIGIKNGNATTINNTAD---GVINVYADDSYAFGGKTKAIINNGEINLLC 1466
Qy 1579 DTGTTAKKLGTLTIKGGQDTNKLTONNI---GVVAGTDGFTVKLAD-----LTN-LN 1629
Db 1467 DSGCDIVAPGTT---GTQNDHNGTADIVDPDAPTETGSIPTPPADNPAPQQLSNVIV 1522
Qy 1630 SVNAGGTKIDDKGVSVFVDSGQAKANTPVLSSANGLDLGGKVISNVGKTGDT-----1681
Db 1523 GTNADG-----SSGTLKANNLVIGDN-----VKVDTGFTSGTADTTVVVDNAF 1565
Qy 1682 DAANVQQLNEVRNLLGLGNA-GNDNADGN-----QVNIADTKDPNSGSS 1725
Db 1566 TGSNIQADNITSTSVVWNAQSQDADGNDVDTWTKNAYADVATDSSVSDVAQALDAGYT 1625
Qy 1726 SNR--TVIKAGTVLGGKNNDEKLTATGVGVGVGDKNAGNDLSNVWVKTKQDKGSKAL 1783
Db 1626 NNELYTSLNVT-----TALNSALKQV-----SGAQATTVFREARVLSNFTM 1669
Qy 1784 LATYNAAGQNTYLTNNPFAEIDRINEQGRIFPHVNDQNEPVGQNRGIDSSASGRHSVA 1843
Db 1670 LA--DAAPQ-----IKDG-----1680
Qy 1844 IGFOAKADGEAAVAIGRQTQ-----AGNQSIAT-----GDNAQATGDSOI 1883
Db 1681 LAFNVVAKGDPRAELGNDTQYDMLALRQTLDTASQNLTLLEYGIARLDGDKGSKTAGDNL 1740
Qy 1884 AIGTGNVAGKHSGAIGDPSVTKADNSVSGNNGNQFTDATQDVFVGNNITVTESNSVA 1943
Db 1741 TGGYSQFGLKHSWAFDEGLANNLSRYDVIN---LDSSRSVAYGVNKNIDADMKQQY 1796
Qy 1944 LGSNSAISAGTHAGTQAKKSDGTAGTTTTAGATGTVKGFAGQATVAGVAVSGASGARRIQ 2003
Db 1797 L-----EPRSEG-AKTFTMWDALKVTPYAGVKFRHTMEDGYKERSAGDF 1840
Qy 2004 NVAAGEVSATSDAVNSQL-YKATQG-IANATHELDRHONENKANAGISSANAMASM 2061
Db 1841 NLSMNSGNETAVDSIVGLKLDYAGDKGMSATATLEGGPNLSYSQSORTASLQGAAG---1896
Qy 2062 PQAVYIGKRSMTVGGIATHNGQGAVALGSLKSLDNG-----QWVFKINGSAD 2107
Db 1897 -QSFVDDQKGGV---NGLATIGV---KYSSNDTALHLDAYOW--KEDGISD 1941

RESULT 8
FHAB BORPE
ID FHAB BORPE STANDARD; PRT; 3590 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
```

```
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DB 15-MAR-2004 (Rel. 43, Last annotation update)
GN Filamentous hemagglutinin.
OS FHAB OR BP1879.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC NCBI_TaxID=520;
OC NCBI_TaxID=520;
(1)
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor.";
RL Mol. Microbiol. 4:787-800(1990).
(2)
SEQUENCE FROM N.A.
STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RC MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Leithwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Rabinowitsch E., Rutter S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Sharp S., Simmonds M., Skelton J., Squares D., Seeger K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
(3)
SEQUENCE OF 1-3261 FROM N.A.
MEDLINE=89202384; PubMed=2539596;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -!- FUNCTION: Evidence for a role in host-cell binding and infection.
CC -!- SUBCELLULAR LOCATION: Surface.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M60351; AAA22974.1; -.
EMBL; BX640416; CA642162.1; -.
PIR; S21010; S21010.
DR InterPro; IPR008619; Fil haemagg.
DR InterPro; IPR008638; Haemagg act.
DR Pfam; PF05594; Fil haemagg; 18.
DR Pfam; PF05860; Haemagg act; 1.
KW Antigen; Hemagglutinin; Complete proteome.
FT CONFLICT 507 508 KQ -> NE (IN REF. 1 AND 3).
FT CONFLICT 1454 1454 A -> P (IN REF. 1 AND 3).
FT CONFLICT 3574 3590 VEDIGKNYRVFYETNK -> SRISAARTGTSSMKPTNR
FT CONFLICT 3574 3590 (IN REF. 1).
SQ SEQUENCE 3590 AA; 367519 MW; C00BD8E22C9DB41D CRC64;
Query Match 4.5%; Score 480.5; DB 1; Length 3590;
Best Local Similarity 19.9%; Pred. No. 3e-10;
Matches 492; Conservative 306; Mismatches 900; Indels 769; Gaps 110;
Qy 9 FNKATGTFMAVAYAKSHSTGG-----GSCATGQGVSVRTL-----SFARIAALAVLVIGAT 60
Db 264 YDHATRTATPIAGARGAAGAYADGTAGAMYGKHTLVSSDSGLGVLQSLSPSA 323
```


Db 2130 ADMRALGHSQLMQRKWKFKAGRGAEIA-----FYPKEQT-VLAAGA 2170
 Qy 1933 NITVTEGNSVALGNSAISAGTHAGTQ--AKKSDGTAGTITTTAGATGTVKGFAGQTAVGA 1990
 Db 2171 GLTILS-NGAIHNGENAAQNRGPEGLKIGAHSAISVSGSPDALRDVGLKELDDIDALLA 2229
 Qy 1991 VSVG-----ASGAERIRIONVAAGEVSAT-----STDVAVGSGQLYKA 2026
 Db 2230 VLVNPHFTRIGAAQTSIADGAAGPALARQARQAPETDGMVDARGLGSADALASLASLDA 2289
 Qy 2027 TOGIANATNELDRIHONENKANAGIS--SAAWASMPQAVIPGRSMVTGGIATHNGGGA 2084
 Db 2290 AQLG-----EVSGR--RNAQVADAGLAGPSAAPAAGAAV-DVPEVTG---DQVDQPV 2338
 Qy 2085 VAVGLSK 2091
 Db 2339 VAVGLEQ 2345

RESULT 9

OMPB RICPR STANDARD; PRT; 1643 AA.
 AC Q53020; Q9ZCWO;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB OR SPAP OR SPA OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Breini;
 RX MEDLINE=91045972; PubMed=2122457;
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
 RA "Characterization of the gene encoding the protective paracrystalline-surface-layer protein of Rickettsia prowazekii: presence of a truncated identical homolog in Rickettsia typhi";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Breini;
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria";
 RL Nature 396:133-140 (1998).
 RN [4]
 RP PARTIAL SEQUENCE.
 RC STRAIN=Breini;
 RX MEDLINE=92114896; PubMed=1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of the S-layer protein antigens of Rickettsia typhi and Rickettsia prowazekii";
 RT Mol. Immunol. 29:95-105 (1992).
 RN [5]
 RP CLEAVAGE SITE.
 RX MEDLINE=92104668; PubMed=1729180;
 RA Hackett T., Messer R., Cieplak W. Jr., Peacock M.G.;

RT "Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein of rickettsiae: identification of an avirulent mutant deficient in processing";
 RL Infect. Immun. 60:159-165 (1992).
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; M37647; AAA26390.1; ALT INIT.
 EMBL; AF161079; AAD42234.1; -
 EMBL; AJ235273; CAAL5140.1; -
 PIR; D71630; D71630.
 DR InterPro; IPR006315; Autotransporter.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRFAMs; TIGR01414; autotrans_barl; 2.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1328 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1329 1643 32 kDa BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREINL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTQEAPLTLGA -> INSRSSSYHLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

Query Match 4.5%; Score 480; DB 1; Length 1643;
 Best Local Similarity 22.2%; Pred. No. 1.2e-10;
 Matches 396; Conservative 222; Mismatches 619; Indels 544; Gaps 97;

Qy 581 TNSNKQIQVGADGITTDINSRPGAGIE-----NTRITRDGIGFANNVTSGLDANKPRL 635
 Db 18 TASTATIVAGFGVAM-----GAAMQYNRTTAAATTFDGIQFQAAAG---ANIPVA 66
 Qy 636 TPTGINAGGKELTNVOSAINPAT-----NGGQIDFMNRLSTANTEKSSAATIKDLY 687
 Db 67 PMSVITA-----NANNPTFTNPNGHLSFLDTANDLAVTINEDT-----TLGFI 113
 Qy 688 NLSQVPLTFAGDTPNVTYKGLKGLKGVKGTITADD-----LTKNIGVADSTD 737
 Db 114 NIAQAQKFF-----NFTVAAAGKILNTIGQITVQEASNTINAQNALTKVHGAAINAWD 167
 Qy 738 ----NSLITVKLAKTSLDLDVNTK-----TLTASDKVTVDGNNVAKLQNGDLTFSKQN 787
 Db 168 LSLGLSGITFAAPSVLEFLNINPTQEPILTLGANSKI-VNNGNGTLNITNGFIQVS-DN 225
 Qy 788 TGATPATNSKTIQVGLKFTDNNGIALDGTITYITKDKVGFQKQDGLSKPYLDKDKLK 847
 Db 226 TFA-----GIKTINID-----DCQGLMFSNTP-----DAANTL-----NLQ 256
 Qy 848 VGEVEITTTNGINAGGKAITGLSN---TLTDATNATTTGHVTLQGLIVDSTDKTRAASI---G 901
 Db 257 VGGNTINFNGIDGTGKLVLSVKNKAATEFNVTGLGNLKG--GIIEINTAAVAGKLISQ 314
 Qy 902 DVLNAGFNLKN--GDAKDFVSTYDVTDFINGNATTAKTAVYDGKASKAVDVNVDG---TTI 958
 Db 315 GAANAVIGTDNGAGRAAGFIVSD-----NGNAATISSGVYAK--NMVIOANAGGQVTFE 368


```

RT elements.";
RL Virology 167:497-506 (1988).
RN [2]
RP REVISIONS.
RX MEDLINE=21342589; PubMed=11448171;
RA Jakob N.J., Mueller K., Bahr U., Darai G.;
RT "Analysis of the first complete DNA sequence of an invertebrate
RT iridovirus: coding strategy of the genome of Chilo iridescent virus.";
RL Virology 286:182-196 (2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF303741; AAK82303.1; ..
DR PIR; A31828; REXFII.
DR PIR; B31828; REXFIJ.
DR PIR; C31828; REXFII.
DR PIR; D31828; REXFIL.
DR PIR; E31828; REXFIM.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 2432 AA; 237370 MW; C493EF862A56F863 CRC64;

Query Match 4.4%; Score 466.5; DB 1; Length 2432;
Best Local Similarity 21.1%; Pred. No. 6.2e-10;
Matches 493; Conservative 316; Mismatches 890; Indels 633; Gaps 122;

QY 13 TGTGMAVAYAKSHSGGSCATGQVGSVRTLSFARIALAVLIGATLNGSAYAGIGIS 72
DB 13 TGTGMAVAYAKSHSGGSCATGQVGSVRTLSFARIALAVLIGATLNGSAYAGIGIS 72
QY 353 TGDYIITD---PPTVGTGAANKAYVDANITPN-ATPTVLGKIQLSGDISGTAIAPVWSP 408
DB 353 TGDYIITD---PPTVGTGAANKAYVDANITPN-ATPTVLGKIQLSGDISGTAIAPVWSP 408
QY 73 EADGKGANARGDKSIAIGDIAQALGSGSIAIGDNKI VHSNNNANIGAKAGNESIAI 132
DB 73 EADGKGANARGDKSIAIGDIAQALGSGSIAIGDNKI VHSNNNANIGAKAGNESIAI 132
QY 409 GAITLSKMANAVSNL-IGSSNTNVTPTNISLGSNLQMTGTLNVNL-TSLSGSFLSLL 466
DB 409 GAITLSKMANAVSNL-IGSSNTNVTPTNISLGSNLQMTGTLNVNL-TSLSGSFLSLL 466
QY 133 GGDVLASGHASIAIGDDLYLKBTQQISELLPIIRGQKALNDIYQIADTNIQKRYRTH 192
DB 133 GGDVLASGHASIAIGDDLYLKBTQQISELLPIIRGQKALNDIYQIADTNIQKRYRTH 192
QY 467 GGTM--SG--NIIIPSGDL-----ISIADAPVSGTSAANKSY--VDSQIIVNATEN 511
DB 467 GGTM--SG--NIIIPSGDL-----ISIADAPVSGTSAANKSY--VDSQIIVNATEN 511
QY 193 AQCHASTAVGAMSYAKGHFSNAPGRTAEGTVSLAVGLTATATAKASSIAVGSNAQAGF 252
DB 193 AQCHASTAVGAMSYAKGHFSNAPGRTAEGTVSLAVGLTATATAKASSIAVGSNAQAGF 252
QY 512 A---TSTVLGKIQLT-----GDLGSSATPP---TVAPGAILTSLKAN---LSPSKLIQS 558
DB 512 A---TSTVLGKIQLT-----GDLGSSATPP---TVAPGAILTSLKAN---LSPSKLIQS 558
QY 253 AATAVGSGTQVNLNRGIALGFGSQVLQKQNDVNAAVRAVAPDDNOPIDNRYKATFKNGA 312
DB 253 AATAVGSGTQVNLNRGIALGFGSQVLQKQNDVNAAVRAVAPDDNOPIDNRYKATFKNGA 312
QY 559 GSTS---SSPANITLGTSLMSGTSL-----NV---VPTFSNP-----TNGTI 596
DB 559 GSTS---SSPANITLGTSLMSGTSL-----NV---VPTFSNP-----TNGTI 596
QY 313 TDVFSIGNSN---GNDISIRKIIIVGAGSADTDVNN---VAQKEAVRLANROITPKGDD 366
DB 313 TDVFSIGNSN---GNDISIRKIIIVGAGSADTDVNN---VAQKEAVRLANROITPKGDD 366
QY 597 SGTAVLGVNGSGGTGNTSLGYV--VNGTAPFTAVTSIPVSNVNGAVQ----- 642
DB 597 SGTAVLGVNGSGGTGNTSLGYV--VNGTAPFTAVTSIPVSNVNGAVQ----- 642
QY 367 SNRVEKGLGKLTITGGAGTSA-----LTDHNGIVV-----QNG----- 401
DB 367 SNRVEKGLGKLTITGGAGTSA-----LTDHNGIVV-----QNG----- 401
QY 643 SVNGVFPDLSCNVTVLGVITGTLAALPPVGPPLVNGDIYVVGSDPTPSNGLTIYST 702
DB 643 SVNGVFPDLSCNVTVLGVITGTLAALPPVGPPLVNGDIYVVGSDPTPSNGLTIYST 702
QY 402 -----DGLKVQLAE-----TLTSLKMTTENLTANE--- 427
DB 402 -----DGLKVQLAE-----TLTSLKMTTENLTANE--- 427
QY 703 TPTNQLWLEISPSFGSLDARYLQLSGGTMSGNIVIPSGNFTILTSLPVNPFD--AANKSYV 760
DB 703 TPTNQLWLEISPSFGSLDARYLQLSGGTMSGNIVIPSGNFTILTSLPVNPFD--AANKSYV 760
QY 428 --KVTVGKRLTDDKLGFTNDMMGIDSKPYLDKDTGIHAGQKQIKTLTAGVVDADAATY 485
DB 428 --KVTVGKRLTDDKLGFTNDMMGIDSKPYLDKDTGIHAGQKQIKTLTAGVVDADAATY 485
QY 761 DVNITPSAITSLLQKQVLSGDLISGV-----ASAPVITTGAILT 798
DB 761 DVNITPSAITSLLQKQVLSGDLISGV-----ASAPVITTGAILT 798
QY 486 GQLKKVNTQAESALQITFTVKVVDKNGDANDSKIIITVGKNNKPDGT--QVNTLKLKENG 543
DB 486 GQLKKVNTQAESALQITFTVKVVDKNGDANDSKIIITVGKNNKPDGT--QVNTLKLKENG 543
QY 799 --TKMANLTSTSSL-----IGSSSTSTSPSQ--LSLGSNLQISGTTLDVNTSSLSG--- 845
DB 799 --TKMANLTSTSSL-----IGSSSTSTSPSQ--LSLGSNLQISGTTLDVNTSSLSG--- 845
QY 544 VDVTTETNGVTVFGLNQNGTLVCGNLTNNDLGSLVKNNTSNKQIQVGADGITPTDISNSK 603
DB 544 VDVTTETNGVTVFGLNQNGTLVCGNLTNNDLGSLVKNNTSNKQIQVGADGITPTDISNSK 603

```

```
QY 1528 VNDANKQGINEDNAFVKGLEKASDNKTKNAAVTVGDLNNAVAQTPLTFAGDGTCTTAKKL 1587
Db 1839 -----SSTFL-----PLAGGTWNGNIIPTGDLISADAPL--VGTSG--ANK- 1877
QY 1588 GETLITKGGQDTNKLNNIIVVAGTGDFTVKLAKDL-----TNLNS--VNAGG---TKI 1638
Db 1878 -----SYDQSIIANATPSAITGIQ-KIQLAGDLGSGTTPASSVIGSAILTKM 1928
QY 1639 DD-KGVFVSSGQAKANTPVL-ANGLDLGGKVIS-----NVG----- 1675
Db 1929 ANLSGNSQIIGSGTSSPNLTLGSGLIQSGTVLSVNSATLTPPATATTIGGIEMGLD 1988
QY 1676 ---KGTKDADAANVQQLNEVRNLLGLN-AGNDNADGNQVIA-----DIKDPNSGSS 1725
Db 1989 LTGSVATAPTIAAGAILAKMANLSGNSQIIGSSSTTSTPTNLTLGSLQISGTVLSVNS 2048
QY 1726 SNRTVIKA-GTVLGG-----KGNNDT-EKLTATGGVQVG--VDRKGNAN--GDLNSVWV 1772
Db 2049 ATLTVPATATTIGGIEMGLDGTGSVATAPTATGATLTKSWANLSGNSQIIGSSSTTST 2108
QY 1773 KTXD-CGSKKALLATYNAGQNTVLTNNPRAIDRINEQGIREFHVDGN--QEPVVOGR 1829
Db 2109 PTNLTLGSLQISGTVLSVNSAT-LTPVPATA---TTGGIEMGLDGTGSVATAPT--- 2161
QY 1830 NGIDSSAGKHSVAIGFOAKADGAAVAIGRQTQAGNQSIAIGDNAQATGDSQIAIGTGN 1889
Db 2162 -----AAG--AIVLAKWANLSGTSQIIGSSSTTSPNISLGLTQMSG-TTILSVNTST 2212
QY 1890 VVAGKHSAGIDPPTVKA-----DNSYSVGNNGNFTDAT-----QTDVFGVGNIT 1935
Db 2213 LMLLVPSSVNGDLATLNASGOVIDSGVSI--NNSGLTSASLWNAAKLAITNSWEAGTN-- 2269
QY 1936 VTENSVALGNSALSAGTHAGTQAK-----KSDGTAGTTTATAGT- 1977
Db 2270 ---PNTTAPTRDPATSSVLYVGTGDTASLWNGSVYISLIGAKVPTSVTRTFTTSGSGF 2326
QY 1978 ---TVKG-FAGQTAVGAVGSCGAERRIQNVAAEVSATSDA-----VNG 2020
Db 2327 QISTINGAFVHYSVISITIGVGTSTGTNL---EVSPTNSATPASHVING 2375

RESULT 12
OMPB_RICCN STANDARD; PRT; 1655 AA.
AC Q9KK3; Q9KK38; Q9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) (Contains: 120 kDa surface-exposed protein [Surface protein
DE antigen] (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB OR RCI085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii."
RL Science 293:2093-2098 (2001).
RN [2]
RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
```

```
RT gene coding the outer-membrane protein rOmpB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455 (2000).
RN [3]
RP SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008659; AAL03623.1; -
DR EMBL; AF123721; AAF34124.1; -
DR EMBL; AF123726; AAF34129.1; -
DR EMBL; AF149110; AAD39533.1; -
DR PIR; E97835; E97835.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_barl; 2.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
FT VARIAT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 353 354 KD -> GH (IN REF. 3).
FT CONFLICT 776 776 F -> S (IN REF. 3).
FT CONFLICT 1159 1159 E -> D (IN REF. 3).
FT CONFLICT 1177 1177 G -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match 4.3%; Score 456.5; DB 1; Length 1655;
Best Local Similarity 22.3%; Pred. No. 9.2e-10;
Matches 388; Conservative 194; Mismatches 663; Indels 494; Gaps 94;

QY 630 ANKPRITPTGINAGGKELTNVQSAINPATNGQLDFMNLSTANTERSGSAATIKDL-YN 688
Db 2 AQKPNFLKLLISAG---LVTASTATIVASFAGSA--MGAAIQQRTTNAVATTVDGVGFD 56
QY 689 LSQVPLTAFAGDTGPNVTKKLGEILKVGKKTADDLTAKNNGVADVSTDSNLTKVLAKTL 748
Db 57 QTAVP-----ANVAVPLNAVITAGVNK-----GITLNPAGSFNGFLNTA 97
QY 749 SLDPAVNTKLTATSDKVT--VDSGNNTAKLQNGDLTFKQNTGATPATNSKTIGVDGLKF 806
Db 98 NNLD-VTVREDTTLGFTTNVNNANHFNLMLNAGKTLTITCGGTTNVQAAATKNNVVA 156
QY 807 TDNNGIALDGTYY--ITKDKVGFQKQDGLSKPYLDKDKLVKGEVETITNGNAGGKA 864
Db 157 QVNNGAIDNNDLQGVGRIDCGAAASTLVFNLANPTTQKAPLILGDNAVIVNGANG---- 212
```

QY 865 ITGLSNTLTDAATNATTGHVTLQGLVDS TDKTRAASIGDVLNAGFNKNNGDAKD--FVST 922
Db 213 -----TLNVTNGP-----KVSSKSPATV-----NVINIGDGGIMFNFTD 247
QY 923 YDVPFINGNATTAKVTV---DGKASKAVYDVNVVDGTTIHLTGA-DGNKNQIGVKVTTTLT 978
Db 248 ADVNLTNLQANGATITFNGTDTGRLVLLSKNAATDENVTGSLGNLKGIIIEPNTVAV 307
QY 979 KTDAGDKKAINSV---NSGDDKAL-----INAKDIADNLNLTAGEIR 1018
Db 308 NGQLKANAGANAAGVGTNNAGRAAGFVVSVDNGKVATIDGQYAKDMVIOQSANAVGQVN 367
QY 1019 -----NTKGTADTALOTFQVK-KVKENG-----DDDDADATITVKGDKAKTNQVNTLKL 1065
Db 368 FRIHVDVGTGDT--TAPKTAASKVAITQNSNFGTTDFGNLAAQIIV-----PNTMTL 417
QY 1066 KGKNGLDIOT--NKDGTVTFFGIN--TOSGLKAGNNTTLNKN-----GLSIRKNTAGNE-- 1113
Db 418 NGNFTGDASNPGNTAGVITFDANGTLASASADANAVATNNITAIEASGAGVQVLSGTHAA 477
QY 1114 QIQVG-----ADG-KVFAKN-----NGVVGAG---IDGTRITRDEIGPAGTNGSL 1156
Db 478 ELRLGNAGSVFKLADGTVINGKVNQNTALVGGALAAAGTITLDGSATITGD--IGNAGGAAL 536
QY 1157 DKSKPHLSKDG---INAGGKKTINIQSGEIAQNSNDAVTGGKIYDLKTELEKNISSTAKT 1213
Db 537 QGIT--LANDATKTLTGGANIIGANGTI---NFQANGTI-----KLSTST--- 578
QY 1214 AQNSL---HEFVSADQGNFTVSNPYSDYTSKTSVDITFAGENGITTKVKNKGVVRVGI 1270
Db 579 -QNNIVDFDLAIAIDQ---TGVDASSLTNAQT---LTINGKIGTVGANNKTLGQFNI 630
QY 1271 DQTKGLTTP-----KLTVGNKN-----GKG-IVIDSQNGQNTITGL 1305
Db 631 GSKSVTLSDGVDAINELVLTGNNGAVQFAHTYLTITTTNAAGGKIIIFNPVNNNTLAT 690
QY 1306 SNTLANVNDKGSVRTTEOGKIKEDKTKRAASIVDVLSGAFNLQNGRAVPFVSVDYTV 1365
Db 691 GTNLGSATNPLAINEPGSKG-----AANVDVTLNVGKGV---NLVATNITTTIDANV 738
QY 1366 -----NFADGNATTAKVYTDTSKTSKVYDVNVVDTTIEVKDKKLGVKT---TTLTST 1416
Db 739 GSFIENAGGTNI VSGTVGGQGNKFNKNTVALD---NGTIV---KPLGNATFNNGTITIAAN 791
QY 1417 GT---GAN---KFALSNOATGALVKASDIVAHLNLTSGDIQTAKGASQANSAGY-VDA 1469
Db 792 STLQIGGNTADFVASADGTG-----IVEFVNTGPTITVTILNKQAPFNALKQITVSG 843
QY 1470 DGNKVIYDSTDNKKYQAKN-DGTVDKTEKVAKDKLVAQAQTP-----DGTLAQMNKVS 1521
Db 844 FGNVVI-----NEIGNAGNYHGAVDTTIAFENSSLGAVVFLPRGPFNDAGNRIPIUTKS 898
QY 1522 VINKEQVNDANKKQGINEDNAFVKGLEKASDNKT---KNAAVTVG---DLNAVAQTPLT 1575
Db 899 TVGNKTAT-----GPDVPSVILVGVDSVIADQVIGDQNNIVGLGLGSDNDIIVNATTL 952
QY 1576 FAGDTGTTAKLGETLITIKGGQTDI-----NKLTD-NNIGVVAGT- 1614
Db 953 YAG-IGTINNQG--TVTLSGGHPNTPGTVYGLGTGIGASKFKQVPTPTTDYNNLGNIAFN 1010
QY 1615 ----DGFTVKLA-----KDLTNLSNVNAGGTKIDDKGVSVFDS-----SGQAKN 1655
Db 1011 ATINDGVTVTTGGIAGIGFGDKITLGSVNGG-----NRFVDDGILLSHSTSMIGTTKAN 1064
QY 1656 TPVLSANGLDLGGKVISNVKGTGKTDDAANVQOLNEVRNLLGLNAGNADNADGNQVNIAD 1715
Db 1065 NGTVTV---LGNAFVGN--GSDTTPVASVR-----FTGSDGGAGLQNGIYS 1106
QY 1716 IKKD---PNSGSSSRNRTVIKAG-TVLGGKGNNDTEKLATGGVGVGVVDKGNANGDLSNVW 1771
Db 1107 QVIDFGTYNLGISNSNVILGGGTGAINGKINRLTNTLTF-----ASG--TSTW 1152

QY 1772 VKTKDGSKKALLATYAAGQNTLYNTNPAEADIRNEQGIREFHVDGNEQEPVQVGRNG 1831
Db 1153 -----GNWTSIETTLTTLA-----NGNIGNIV----- 1173
QY 1832 IDSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNOSIAIGDNAQA--TGDQSIATGTGN 1889
Db 1174 -----ILEGAVNA-----TTTGTTTIKVQDNANANFSQTQTYTLIQG- 1211
QY 1890 VVAGKHSIGAIDPSTV-----KADNSYSGVNNNOFTDATQTDVF----- 1928
Db 1212 --GARENGTLGPFNVVTGSRNFRVNYGLIRAAANQDYVITRTNNAENVVNTDINANSFPGA 1269
QY 1929 -GVGNNT--VTESNSVA-----LGSNSAISAG-----THAGTQAKKSDGTA 1967
Db 1270 PGVGQNVTTTFVNATNTAAVNNLLAKNSANSANFVGAIVTDTSSAATNQLDVAK-DIOA 1328
QY 1968 GTTTTAGA-----TGTVKGFAGQTVAGSVGASGAERRIQNVAAAG-EVSATSTDA--- 2017
Db 1329 QLGNRGLALRYLGTPTETAEMAGPEA-GAIPAAVAAGDEAVDNVAYGIWAKPFYVTDHQSK 1387
QY 2018 VNGSOLYKA--TOGIANATNELDHRHONENKANAGISSAMAMASMP---QAYIPGRSMVT 2073
Db 1388 KGLAGYKAKTGWIGLDTL-----ANDNLM---IGAAIGITKTIDIKHQDYKKGDKTDV 1439
QY 2074 GG-----IATHNGQGA VAVGLSKLSLSDNGOMVFEKINGSADTQGHVGAAGVAGF 2121
Db 1440 NQFSFSLYCAQOLVKNFFAQSIAFSLNOVKNSQRYF-----FDANGNMSQIAAGHY 1493

RESULT 13

P3P_LACLC
ID_P3P_LACLC STANDARD; PRT; 1902 AA.
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-
DE associated serine proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN (1)_LACLC
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase."
RL J. Biol. Chem. 264:13579-13585(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on casein,
CC although it has been shown to hydrolyze hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; J04962; AAA03533.1; ALT_SEQ.
CC HSP; P00782; 2SET.

DR MEROPS; S08.019; --
 DR InterPro; IPR001893; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PSS0847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
 KW Signal; Plasmid.
 FT SIGNAL 1 33
 FT PROPEP 34 187
 FT CHAIN 188 1870
 FT PROPEP 1871 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT SITE 1867 1871
 FT MOD_RBS 1870 1870
 SQ SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;
 Query Match 4.2%; Score 444.5; DB 1; Length 1902;
 Best Local Similarity 20.8%; Pred. No. 3e-09;
 Matches 443; Conservative 219; Mismatches 753; Indels 715; Gaps 100;
 370 RVEKGLGKLTITIGGAOTSALTDRNIGVQVQNGDLKQV-----LAETLT--SLRWVTEN 422
 DB 3 RKKKGL--SILLAGTVALGALAVLPVGEIQAKAISQOTKSSSLANTVTAATAQAQATDT 60
 423 LTANEKVTVGKRLTTDKIGFTNDMNGIDSKPYLD----- 458
 DB 61 TAAATTNQAIA--TQLAAKGIDY--NKLNVKQODIYVDVIVQMSAAPASENGILRTDYSSTA 118
 459 -----KDTGIHAGGQKIITLKTAGVVDVDDAANYG-----ESGYVYVNGFSTKRVVVDIPKLKQI 172
 DB 119 EIQOETNKVIAAQAASVKAAVEQVQQTAG----- 494
 495 AESALQFTTQVKKV-----DKNGNDANDSKIIITVGKNNKPDGTQVNTLKLKGBGVDDVT--- 547
 DB 173 A--GVKIVTLAKVYPTTDAKANSMAVQAVMSNYKYKGEIVSVSI-----DSGIDPETHK 226
 548 -----TET-----NGTVTFGLN--QNN----- 562
 DB 227 MRLSDDDKDVKLKSDVEKFTDTVKHGRYFNSKVPYGFNYADNDNTITDDKVDQHGHHVA 286
 563 GLTVGNSTLNDGLSVKNTNSNKOI--OYQADGITTDTISNKPAGAGIENTTITRD-- 617
 DB 287 GIIGANGTGDDPAKSVVGVAPEAQLAMKVFNSDTSAKTGSATVVSALIEDSAKIGADVL 346
 618 GIGFANNTGSLDANKPRILPTGGINAGGKELTNVQSAINPATNGGQLDFMNR----- 668
 DB 347 NMSLGSNSGNTLEDPELAAVQ--NANESGTAAVISAGNSGTSAGTEGVNKKYVGLQDNE 405
 669 -----LSTANTEKSGSAATIKDLNLSQVPLT-----FAG----- 698
 DB 406 MVGSPGTSRGATTVASAENTDVIITQAVTITDGTGLQLGPETIQLSHDFTGSPQKKFYI 465
 699 --DTGPNVTK-----KLGEILKVKGGKTTADDLTK-----NNIGVVAADSTDNSLT- 741
 DB 466 VKDASGNLSKALADYTADAKGKTAIVKRGFSFDDKQYAAAGAGLIIIVNTDGTATP 525
 742 ---VKLAKTLS--DLDAVNTKTLTASDKVTVDSGNN-----TAKLQNGDLTFSKQN--T 788
 DB 526 MTSIALTTTPTFTGLSSVTGQKLV--DWVTAHPDPSLSGVKITLAMLPMQKYTEDKMSDFT 583
 789 GATPATNSKTIQVGLKFTDNNGTALDGTITTKDKYGFQKQDGL----- 834
 DB 584 SYGPVSN-----LSFKPDI--TAPGGNIWSTQNNNGYTNMSTGSMASPPFIAGSQLLK 634

QY 835 ---DKSKPYLDKDKLVG-----EVEITT-----NGINAGGKAITGLSNTLTDATNA 878
 DB 635 QALNNKNPFYAYYKQLKGTALTDFLXTVEMTAQPINLIN-----YNNVIVSPRRQ 686
 QY 879 TTGHTVQLGIVDSTDKTRAASIGDVLNAGFNLKNG-----DAKDFVSTYDTVDVFNAT 934
 DB 687 GAGLVDVKAALDALEKNPSTVVAE-----NGYPAVELKDFSTDKTFLTFNRT 736
 QY 935 TAKVY--DGA--SKVAYDVNVDTTI----- 958
 DB 737 THELTVMDSNTDTNAVYTSATDPNSGVLYDKKIDGAAIKAGSNITVPAGKTAIEFTLS 796
 QY 959 ---HLTGADGNKQI-----GVK-----TT 975
 DB 797 LPKSPDOQFVEGFLNFKGSGSRNLNPFMGFFGDMNDGKIVDSLNGITVSPAGNGFCTV 856
 QY 976 TLTK-----TDAG-----DKAINFSVNSGDDKALIN----- 1002
 DB 857 PLLKNKNTGTQYGGMVTADGNKTVDDQAIAP-----SSDKNALYNDISMKYLLRNISNV 913
 QY 1003 AKDIAD--NLNTLAGEIRTK-----CTADTAL 1028
 DB 914 QVDILDGQGNKVTTLSSSTRKRTYYNAHSQQYIYINAPAWDGYTDQRDGNITKADGGS 973
 QY 1029 QTFQVKKVKGDD-----DNDADTIT--VGDAKTQNVNT---LKLKKG--NGL 1071
 DB 974 YTRISGVPEGDKRQVDFVPFKLDKAPTVRHVALSAKTENGKTQYLLTAEEKDLSGL 1033
 QY 1072 DIQ-----TNKDGTVTFGINTOSGLKAGNNTLLNNGLSIKNAGNEQIQVGADG 1121
 DB 1034 DATKSVKTEINEVNTLDATFTDAGTADGY-----TKIETPLSEQAQA--- 1077
 QY 1122 VKFAKNVNGVAGIDGTTRITRDEIGFAGTNGSLDKSKHLSKDG--INAG--KKITN 1177
 DB 1078 ---LNG-----DNSAELYLTDNASNATDQASVQKPGSTSFDLIVNGGGIPDKLSS 1126
 QY 1178 IQSGEIAQNSDAVTTGGKIYDLKTELENKISSAKTAQNSLHESPSVA--DEGNNFTVSNP 1236
 DB 1127 TTTGYEANTQ-----GGGYTFSGTYPAADVDTYDAQGGKHDLLNTYDAATNSFTASMP 1181
 QY 1237 YSSYDTSKTSDFVITFAGENG--TTKVNKGV-----VRVGIQDTKGLT----- 1277
 DB 1182 VTNADYAAQVDLVADKAHTQLLKHFTKVRILMAPTFTDLKFNNGSDQTSATIKVGTGVS 1241
 QY 1278 --TPKLTVGNNGKGIVIDSQNGQNTITGLS---NTLANVTNDKGSVRTTEQKIIKDED 1332
 DB 1242 ADTKTVNVGHTVA---ALDAQHHFSVDVPVNYGNTIKVTAIDKQGNNTTTEQKLTSSYD 1298
 QY 1333 KTRAASIVDVLSAGFNLQNGEAVDFVSTYDT--VNFADG--NATKAVTYDDTSKTSKV 1389
 DB 1299 P-----DMLKKS-----TFDQGVKFGTNNKFNATSAKF--YDPKTGIATIT 1337
 QY 1390 YDVNVDDTTIEVKDKKLGVK-----TTTLTSTGTGANKFAL-----SNQATGALVKAS 1438
 DB 1338 GKVKHPTTLTQVDGKQIPIKDDLTFSFTDLGLTGKQFPFGVVGWDTTQNKTFQEALSFIL 1397
 QY 1439 DIVAHLNTLSGDIQTAKGASQANSSAGYVDADGNKVIYDSTDNKYVQAKNDGTVDKTEV 1498
 DB 1398 DAVA--PTLSDSST-----DAPVYTDNPNFQITGATDINAQYLS----- 1435
 QY 1499 AKDKLVAQAQTPDGTTLAQMNVKSVINK--EQVNDANKQKQ---INEDNAFVKG---LEKAA 1551
 DB 1436 -----LSINGSSVASQVEDININSGKPGHMAIDQPVKLLGKKNLTVAV 1479
 QY 1552 SDNKTNAATVVDLNAVAQTPLTFAGDTGTTAKKLGELTILKGGQDQTNKLTNNIGVV 1611
 DB 1480 TDSBN---TTTKNITVYEPFKTLAAPTTPS-----TTEPAQVIT--LTAN----- 1522
 QY 1612 AGTDGFTVKLAKDLTLNLSVNAAGTKIDD---KGVSFVDSGQAKANTPVLISANGLDLGG 1668
 DB 1523 AAATGETVQVSAD-----GGKTYQDVPAAGVT--ITANGTFKSTDLGYGNESPAVD 1572


```
QY 813 A-----LDGTYITKDKVGFQKQDGLSKPKYLDKOKLKVGEVEIT-----NGINAGGKA 864
Db 646 AYYQLKGTAL-----TDFLKT--VEMTAQPININ-----675
QY 865 ITGLSNTLTDATNATTGHVTLQGLGIVDSTDKTRAASIGDVLNAGFNKNG-----DAKDFV 920
Db 676 ---YNNVIVSPRQAGLVDAKAIDALEKNPSITVAE-----NGYPAVELKDF 722
QY 921 STYDVTDFINGNATTAKVY-----DGKA-----SKVADVNVDGTTI-----958
Db 723 STDKTEKLTFTNSTHETLTQMSNTDNTNAVYTSATDPSNGVLVDKIDGAAGAAIKAGSNIT 782
QY 959 -----HLTGADG-----965
Db 783 VPAGKTAQIEFTLSLPKSFQOQFVEGLNFKGSDGSRLLNLYPMGFFGDWNDGKIVDSL 842
QY 966 -----NKNQIGVKTTLTKTDAG-----DKAINFSVNSGDDKALIN 1002
Db 843 GITVSPAGGNFGTVPLLTNKN-TGTQYVGMVTDADGNQTVDDQAIAP-----SSDKNALYN 898
QY 1003 -----AKDIAD-----NLTLAGETRNK-----1021
Db 899 DISMKYLLRNISNVQVDILDGQGNKVTTLSSSTNLTKTYNAHSQYIYINAPAWDGY 958
QY 1022 -----GTADTALQTFQVKKVKGDD-----DNDADTIT-VGKDAKTNOVNT 1062
Db 959 YDQRDGNIKTADGSGYITRISGVPEGGDKRQVDFVPFKLSKAPTVRHVALSAKTENGKT 1018
QY 1063 ---LKLKKG---NGLDIQ-----TNKDGTVTFGINTQSGLKAGNNTLNNGLSI 1106
Db 1019 QYLYLTAERAKODLSGLDAFKSVKTAINEVTNLDAFTDAGTADGY-----TKI 1066
QY 1107 KNTAGNTOIQVAGDQVGFPAKVNNGVVGAGIDGTRITRIDEIGFAGTNGSLDKSPHLSKD 1166
Db 1067 ETPLSDEQAQ-----LNG-----DNSAELYLTDNASNATDQASVQKPGSTF 1111
QY 1167 G--INAGG--KKITNIQSGELAQNSNDVATGKGYDLKTELENIKISSAKTAQNSLHEFS 1222
Db 1112 DLIVNGGGIPKISSTTYGYEANTQ-----GGGYTFSGTYPAAVDGYTYNAQKKHDLN 1166
QY 1223 VA-DEQGNNTVSNFYSDYTSKTSVDITFAGENGI-----TTKVNKGV-----VRGI 1270
Db 1167 TTYDAATNSFTASMPVTNADYAAQVDLYADRAHTQLLKHFDTKVRLTAPTFTDLKFNNGS 1226
QY 1271 DQTKGLT-----TPKLTVGNNGKGVIDSQ-----QONTITGLSNTLA 1310
Db 1227 DQTSBATIKVTGTVSADTKTVNVDGTV-----ALDAQHHSFVDVPVNYGDNITIK-----V 1277
QY 1311 NVTDNKGSRVTTEQGIKIKDEKTRAASIVDVLGAGFNLCNGEAVDFVSTYDVTNVPADG 1370
Db 1278 TATDEGNT-TTEQKTISSYDP-----DMLK-----NSVTFDQGVTFGANEF--- 1319
QY 1371 NATTAKVTDYDTSKTSKVYVQVNDVDDTIEVKDKKLGVK-----TTTLTSTGTGANKPAL 1425
Db 1320 NATSAKF-YDPKGTGIATIGVKHPTTTLQVDGKQIPKIDDLTFSFTDLGLTGLGKPGFV 1378
QY 1426 -----SNQATGDALVKASDIVAHLNTLSGLDIQTAAGKASQANSAGYVDADGNKVIYDST 1479
Db 1379 VVGDTTQNKTFQEALTFLDVA--PTLSLESST-----DAPVYTNDENFQITGTAT 1428
QY 1480 DN-KYVQAKNDGT-----VDKTEBAVDKLVAAQATPDGTLAQNNAKSVINKEQVNDAN 1532
Db 1429 DNAQVLSLSINGSSVASQYVDININSKPGHWAIDQ-FVKLEGNKVLTV-----AVTD-- 1481
QY 1533 KQGINEONAFVKGLEKASDNKTKNAAVTVGDLNANVAQTPLTFAGDGTGTTAKLUGETILT 1592
Db 1482 -----SEDNTTKNTIVVYEPKTL-AAPTV-----TP-----STTEPAKTVTILT 1520
QY 1593 IKGQGTDTNKLTDNIGVAGTGTGTVKLAKDLTLNLSNAGTKIDD-----KGVSFVDSS 1649
Db 1521 ANSAAT-----GETVQYSAD-----GGKTYQDVPAAGVT-VTAN 1553
```

```
QY 1650 GQAKANTPVLSANGLDLGGKVISNVGKGTXTDAAVNVQOL-NEVRNLLGLGNAGNDADG 1708
Db 1554 GTEFKSTDLYGNESPAVDYVVTNI-----KADDDPAQLQAAKQBELTNL-----1596
QY 1709 NQVNIADIKDPNSGSSNRTVIKAGTVLGGKGNNDTEKLTAG--GVQVGVDKDGNGAND 1766
Db 1597 ---IASAKTSLASGKYDDAIT---TALAAA---TQKATDALTQTNASVDSLTGANRD 1644
QY 1767 LSNVWVKTKDQSGKALLATYNAAGQNTYLTNNPAEAIIDINEQIGRFFHFVNDGNQBPVV 1826
Db 1645 LQT-----AINQLAALKPADKKTSL-----NQLOQSV 1671
QY 1827 QGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNQSTAIAGNAQATGQOSTAIG 1886
Db 1672 KAALGTDLGNQTDPSGTCTFTTAAALDLVA-----QAQAGTQT---DDHQAT---LAKV 1719
QY 1887 TGNVAVAGKSGAIGDPSSTVKADNSYSVGNNNQFTDATCTDVFVGNNITVTESNSVALGS 1946
Db 1720 LDAVLAKLABG-----IKAAFTAEEVGNK---DAATGKTWYADIADTLTSGQASADAS 1769
QY 1947 N-----SAISAGTHAGTQAKKSDGTAGTTTTAGATGT---VKGPAGTAVGAV 1991
Db 1770 DKLAHLQALQSLKTKVAAAEAAKTVGKGDTTGTSDKGGQGTTPAPAPGDIGDKGDEG 1829
QY 1992 SVGASGERRIQNVAAEVSATST---DAVNSOLYKATQGIANATWELDHR 2040
Db 1830 SQPSSGG-----NIPTNPATTTSTTDDTTDRNGQLTSGKALPKTGETTER 1876
```

RESULT 15
P2P LACLC
ID P2P LACLC STANDARD; PRT; 1902 AA.
AC P15293;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-associated serine proteinase) (LP151).
GN PRT.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pLP763.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=89313288; PubMed=2501630;
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashina A.;
RT "Molecular characterization of a cell wall-associated proteinase gene from Streptococcus lactis NCDO763.";
RL Mol. Microbiol. 3:359-369(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyze hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X14130; CAA32350.1; --
DR PIR; S06997; S06997.

DR HSP; P00782; 2SBT.
 DR MEROPS; S08.019; --
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_s8.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_s8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRAME; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
 KW Signal; Plasmid.
 FT SIGNAL 1 33 POTENTIAL.
 FT PROPEP 34 187 POTENTIAL.
 FT CHAIN 188 1870 PII-TYPE PROTEINASE.
 FT PROPEP 1871 1902 REMOVED BY SORTASE (POTENTIAL).
 FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 1867 1871 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1870 1870 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1902 AA; 200139 MW; 488D8844D88CDF7 CRC64;
 Query Match 4.1%; Score 437.5; DB 1; Length 1902;
 Best Local Similarity 21.2%; Pred. No. 5.4e-09;
 Matches 457; Conservative 201; Mismatches 728; Indels 773; Gaps 105;
 Db 370 RVEGLGKLTITGGAQTSALTDHNIQVVGQDGLKVQ-----LAETLT--SLKMTYEN 422
 3 RKKGU--SILAGTVALGALAVLPVEIQAKAASQQTGKSSLANVTAAAKQAATDT 60
 423 LPANEKVTVGKRLTDDTKGFTNDMNGIDESPFLD----- 458
 61 TAATTNQAIA-TQLAAKGIDY-NKLNKVOQDIYVDIVVQMSAAPASENGTLRTDYSSA 118
 459 -----KDTGIHAGGQKITKLTAGVDDDAATYG-----OLKKVNOT 494
 119 EIQQETNKVIAQAASVKAAVEQVTOQTAG-----ESYGVVNGFSTKRVVVDIPKLKQI 172
 495 ASSALQTFVKKV-----DANGDANDSKLITVGKNNKPDGTQVNTLKLKGENGVDT-- 547
 173 A-GVKTVLAKVYPTDAKASMANVQAVWSNYKKGEGTVSVI-----DSGIDPTHKD 226
 548 -----TET-----NGTVTFGLN-QNNGLTVGNSTLNN-DGLSVK 579
 227 MRLSDDKVKLTKSDVEKFTDTAKHGRYFNKVPYGVFNVDNNDITDDTVDEQGHMVA 286
 580 NTNSNKQIQVGADG-----ITFTDISNKPQA-----GIENNT 612
 287 GI-----IGANGTGDDPAKSVGVGVAPEAQLAMKVFTNSDTSATTSATLVSAIEDSA 339
 613 RITRD--GIGFANNTGSLDANKPRLTPTGNGAGBELTNVQSAINPATNGGQLDFNR-- 668
 340 KIGADVILNLSGDSGNQTLDPDLAAVQ-NANESGTAAVISAGNSGTSGSGATEGVNKDY 398
 669 -----LSTANTEKSGSAATIKDLYNLSQVPLT-----FAG-- 698
 399 YGLQDNEMVGTPTSGATTVASAENTDVIQTAVITDGTGLQGPETITQLSSNDPTGSF 458
 699 -----DTGPNVTK-----KLGEILKVGKGTATDDLTK-----N 727
 459 DQKFFVVKDASGNLSKGVADYTDADAKGIAIVKRGELTFADKQYQAAGAAGLIIVN 518
 728 NLGV-----VADSTNSLTVKLAKTL----- 748
 519 NDGTATPVTSMAITFTTFFGLSSVTGQKLVDMWAAHPDDSLGVKIALTLVFNQKYTEDK 578
 749 -SDLDVAVN-TKTLTASDKVTVDSGNNTAKLQNGDLTFFSKQNTGATP-ATNSKTIQVDGLK 805
 Db 579 MSDFTSYGVPVSNLSFKPDITAPGNIWSTQNNGNYTNMSGTSMASPFIAQSQALLKQALN 638
 806 FIDNNGIA-----LDGTYITTKQKVGPAKQDGSLSKSPYLDKDKLVKGEVIT--NG 857
 639 NKNPFYAYYKQLKGTAL-----TDFLKT--VEMNTAQPIND 673
 858 INAGGKAITGLSNTLTDAFNATTGHVTLQGIIVDSTDKTRAASIGDVLNAGFNLKNG-- 914
 674 IN-----YNNVIVSPPRQAGLVVDVKAADALEKNEFSTVVAE-----NGYPA 715
 915 -DAKDFVSTYDVFINGNATTAKTY-----DGKA-----SKVAYDVNVDTGTI 958
 716 VELKDFSTDKTKFLTFTNRTTHELTYQMSDNTDINAVVTSATDPNSGVLYDKKIDGAAL 775
 959 -----HLTGADG----- 965
 776 KAGSNITVPAGKTAQIEFTLSLPKSPDQOQFVEGFLNFKGSDGSRNLNLPYMGFFGQDNDG 835
 966 -----NKNQIGVKVTKTTLTKTDAG-----DKAINFSVNSG 995
 836 KIVDSLNGITYSPAGNFGTVPLLTNNK-TGYGGVMVTDAGNQTVDQALAF--SS 891
 996 DDKALIN-----AKDIAD--NLNTLAGEIRNKK----- 1021
 892 DKNALYNDISMKVYLLRNISNVQVDILDGQGNKVTLSSTNRKTKTYNAHSQQYIYHA 951
 1022 -----GTADTALQTFQVKKVKGENDD-----DNADIT-VGKDA 1055
 952 PAWDGTYDQDRDGNKTADDGSYTYRISGVEGGKROQVDFVFPFKLSDSKAPTVRHVALSA 1011
 1056 KTNQVNT--LKLKGGK--NGLDIQ-----TNKDGTVTFGINTQSLGKAGNNTTL 1099
 1012 KTENGKTQYLLTAEBKDDLSGLDASKVKTAINEVNLDATFTDAGTTADGY----- 1063
 1100 NNGLSIRKNTAGNEQIQVGADGVKFAKVNNGVVGAGIDGTTTRITRDEIGFAGTNGSLDKS 1159
 1064 ---TKIETPLSDEQAQA-----LGNG-----DNSAELYLTDNASNATDQDASVQ 1104
 1160 KPHLKSDG--INAGG--KKITNIQSBEIAQNSNDAVTGKIIYDLKTELENKISSTAKTAQ 1215
 1105 KPGSTSFOLIVNGGGIPDKISSTTTGYEANTQ-----GGGTYTFSGTYPAADVDTYDAQ 1159
 1216 NSLHSEFSA-DROGNNFTVSNPYSSYDTSKTSDDVITFAGENGI---TTKVNKGV--- 1265
 1160 GKHLNLTYYDAATNSFTASMPVTHADYAAQVDLYADKAHTQLLKHFDTKVRLTAPTFTD 1219
 1266 --VRVGIDQTKGLT-----TPKLTGVNNGKGVVIDSQN-----GONTIT 1303
 1220 LKFNNGSDQTSKATIKVTGTVSADTKTVNVGDTVA--ALDAQHHFSDVPVNYGNTIK 1276
 1304 GLSNTLANVNDKGSVRITTEQKIIKDEDKTRAASIVDLVSAGFNLQNGEAVDFVSTVD 1363
 1277 -----VTATDEGNT-TTEQKTIITSSYDP-----DMLK-----NSVTFDQGVTFG 1315
 1364 TVNEADGNATTAKVYDDTSKTSKVYDVNVDDTTIEVKDKKLGVK-----TTTLTSTGT 1418
 1316 ANEF--NAISAKF-YDPKGTGATITGKVHHTFTLQVQDKQIPIKDDITFSLDLGLYL 1371
 1419 GANKFAL-----SNQATGDALVKASDIIVAHNLTLSDGIQTAKGASQANSAGYVDADGN 1472
 1372 GQKPFVVVGVDITQNTKTFEALTFILDVA--PTLSLSDST-----DAPVYTNPNF 1421
 1473 KVIYDOSTN-KYYQAKNDGT-----VDKTEVAKDKLVQAQOTPDGTLAQNKKSVINK 1525
 1422 QITGATDNAQYLSLSINGSSVASQYVDININSGKPGHMAIDQ-PVKLLEGKXVLT--- 1477
 1526 EOVNDANKKQGINEONAFVKGLEKAASDNKTKNAATVVDGLNVAQTPLTTFAGDTCTAK 1585
 1478 -AVTD-----SEDNTTTKNTVYEPKKTLL-AAPT-----TP-----STTEP 1513
 1586 KLGETLTIKGGQTDNKLTDNNIGVAGTGTGFKLAKDLTLNLSVNGAGTKIDD---KG 1642
 1514 AKTVLTANSAT-----GETVQVSAD-----GGKTYQDVPAAG 1547

```

QY 1643 VSFVSSGOAKANTPVL SANGLDLGGKVISNVGKGT KDTDAANVOQL-NEVRNLLGLGNA 1701
Db 1548 VT-VTANGTFKFKSTDIYGNESPAVDYVTVNI-----KADDPAQLOAAKQELTNL----- 1596
QY 1702 GNDNADGNQVNIADIKKDPNSGSSNNTVIKAGTVLGGKGNNDTEKLATG--GVQGVGDK 1759
Db 1597 -----IASAKTLSASGKYDDATT-----TALAAA-----TQKAQTALDQTNASVDS 1637
QY 1760 DGNANGDLSNVWVKTKDGSKKALLATYNAAAGQTNLYLTNNPAEAI DRINEQGI RFFPHVD 1819
Db 1638 LTGANRDLQT-----AINQLAAKLPADKKTSL--NQLOSVKAALETDL----- 1679
QY 1820 GNOBPVVQGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTOAGNQSI AIGDNAQATG 1879
Db 1680 GNO-----TDSSTGK-----TFTAALDDLVA-----QAQAGTQT---DDQLQAT- 1715
QY 1880 DQSI AITGNVAVAGKHSAGIDPSTVKADNSYSVGNNNQFTDATQTDVFGVGNNTVTES 1939
Db 1716 ---LAKVLDAVLAKLAEG-----IKAATPAEVGNAK---DAATGKTYADIADTLTSG 1762
QY 1940 NSVALGSN-----SAISAGTHAGTQAKSDGTAGTTTGTAGATGT---VKGFAG 1984
Db 1763 QASADASDKLAHLQALOSLTKVAAA VEAATVKGKGGTGTSDKGGGGQGT PAPTGDIG 1822
QY 1985 QTAVGAVSVGASGAERRIQNVAGEVSATST---DAVNGSOLYKATOGIANATNELDHR 2040
Db 1823 KDKGDEGSPSSGG-----NIPNTPATTSTSTDDTTDRNGQLTSGKALPKGTGETTER 1876

```

Search completed: September 21, 2004, 23:43:46
Job time : 66.4567 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:33:49 ; Search time 70.2522 Seconds
(without alignments)
2905.511 Million cell updates/sec

Title: US-09-813-214A-9

Perfect score: 10708

Sequence: 1 MNHIYKVIENKATGTMAVA.....NGSADTQGHVGAAGAGPHF 2122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1246.5	11.6	2059	D82671	surface protein XP
2	986	9.2	1588	A86036	probable adhesin Z
3	986	9.2	1588	H91188	probable adhesin E
4	728	6.8	1190	A82615	surface protein XP
5	670	6.3	1107	AC0976	probable autotrans
6	669	6.2	3705	AD0123	probable autotrans
7	651.5	6.1	2340	B71704	cell surface antig
8	645.5	6.0	5291	F90696	hypothetical prote
9	634	5.9	5188	B85547	probable RTX fami
10	630.5	5.9	2449	A41477	190K surface antig
11	622	5.8	6713	B89921	hypothetical prote
12	598	5.6	3013	AB0480	probable invasin Y
13	590.5	5.5	2271	F90073	hypothetical prote
14	568.5	5.3	2554	AB3528	hypothetical prote
15	550.5	5.1	2021	A97859	extracellular seri
16	549.5	5.1	2535	AC0304	190-KDa cell surfa
17	548	5.1	4919	T31105	probable hemolysin
18	546.5	5.1	2020	C48399	hypothetical prote
19	541	5.1	2660	E85822	ABC-type transport
20	539.5	5.0	2703	H81193	probable invasin Z
21	536.5	5.0	4152	T31102	hemagglutinin/hemo
22	534	5.0	3624	A08835	filamentous hemag
23	533	5.0	658	AH0110	large repetitive p
24	530	4.9	3890	C89921	probable surface p
25	529.5	4.9	1910	AF0394	hypothetical prote
26	527.5	4.9	5627	C83339	probable adhesin h
27	521	4.9	3295	AB0074	hypothetical prote
28	519.5	4.9	2468	A83412	hypothetical prote
29	517.5	4.8	4936	AH2515	hypothetical prote

30 517 4.8 2481 2 D90011 FmB protein [impo
31 517 4.8 3029 2 S76109 hypothetical prote
32 513 4.8 2514 2 F81045 hemagglutinin/hemo
33 508 4.7 1536 2 A43855 high-molecular-wei
34 508 4.7 2383 2 D64962 probable membrane
35 500.5 4.7 1651 2 JCI1340 outer membrane pro
36 489.5 4.6 1477 2 B43855 high-molecular-wei
37 487 4.5 2479 2 F87386 conserved hypothe
38 483 4.5 1999 2 AB2018 hypothetical prote
39 480 4.5 1643 2 D71630 outer membrane pro
40 478.5 4.5 3591 1 S21010 filamentous hemag
41 477.5 4.5 2055 2 T31110 extracellular matr
42 477.5 4.5 3535 2 E83641 probable hemagglut
43 476 4.4 2893 2 A64556 toxin-like outer m
44 474.5 4.4 1577 2 A35140 hemolysin A precur
45 467 4.4 1487 2 AG2560 hypothetical prote

ALIGNMENTS

RESULT 1

D82671

surface protein XF1529 [imported] - Xylella fastidiosa (strain 9a5c)

C;Species: Xylella fastidiosa

C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C;Accession: D82671

R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequer
Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A;Reference number: A82515; PMID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: D82671

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2059 <SIM>

A;Cross-references: GB:AE003982; GB:AE003849; NID:g9106554; PIDN:AAF84338.1; GSPDB:GN001

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B
Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carriaro, D.M.; Carreir, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranas, E.E.; Laigr
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.B.; Marques, M.V.; Martins, E

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF1529

Query Match

Best Local Similarity 11.6%; Score 1246.5; DB 2; Length 2059;

Matches 578; Conservative 344; Mismatches 780; Indels 681; Gaps 108;

QY	62	NGSAYAGTIGSEADGG-----KGGANARGDKSIAT-----GDIAQA	97
Db	36	NAQVINDGM---DGGCORIVNGSAGVTRVATQCSDEPMTVTGYAFFGSPGTAAEQ	92
QY	98	LG-SQSIAIGNKIVHNSNNNANIKAKAGNSTAIG-GDVL-----ASGHASIAIGSDDL	151
Db	93	QGASRNLTJGGSLLV---NSQGVGVNDVLNKTYISIRNGSVITMTVAGTNAIGS---	145
QY	152	YLKKTVOQISSELLPIIRGQXALNDIYQLADTNLQKRYRTHAQGHASTAVGA--MSYAKG	209
Db	146	-----AQSSAADALKASLTKASG	164
QY	210	HFSNAGFTRATAGTYSIAVGLTATA-KAASSIAVGSNAQALGFAATVAGSGTQVNLNRG	268

165 ARAIAICAKASADGVDTVLGSAGTGTAGTASSIAIGLASAVN-GAVAVGGGALVTVPDG 223
 269 -IALGFGSOVLQKDNVNAANRAYAPDDNQPIDNRYKATFKNGATDVFGISGNXDSI 327
 224 AVALGLNSVA-----STGKGLSGYDP-KTKTTSTDASAAWKSTLAAV-SIGDVSSTNLK 275
 328 RRKIINVAGSADTDANVAOLKEAVLANR--QITEFGDDSNRVE-----KGLKCTL 379
 276 TROLGLAGTSTNDTANVAQLKVVDIEASRGWMLTASGANGVAPGSSVDLKNTKNL 335
 380 TIT-----GCAQTSALTDHNGVQVQGD-----GLKVQLAET---LTSLKMYTTE 421
 336 TITKAIGSNDVQNLNKDVKTTLAVGDALLNTDGLAGTGDVSLTSTGLAITDGPVATAS 395
 422 NLTRANEKV-----TVGKTRLTTPDKTG-----FTNDMNGIDES-----453
 396 GIDAGSKVISHVAAVASET--STDAVNGSOLNAVQVQASQPVFTTNEGAVKRESLQGSV 453
 454 -----KPYLDKDTG-IHA-----GGOKITKLTAGVVDDD 481
 454 VISGESSTAGTSGNLSKVDEAAGRIHLQADSPKPGNVVINGGKISGVITAGTEED 513
 482 AATYGOLKKVNOAESAALQFTTVKVKVDKNGNDANDSKIITVGKXKPKDGTQVNTLKLGE 541
 514 AVNFSQLKSI:STAVD---QGWTLTASGNG-----SKVASGG-----TVDLKNT 554
 542 NGVDVTTETNG---TVTFGLNQN---NGLTVGNSTLNDGLSVKNTNSKQIOVGADGIT 595
 555 DG-NLATISKSGSDSDVFNLSKDFKVDGMTSGTTVWVNDGVK-----GSDVALGTTGLT 608
 596 FTDISNPKGAGIENTTIRITRDGIGFANNTSGSLDANKPRLTPTGINAGGKELTNVQSAI- 654
 609 ITD-----GPAVTASGIDAGSKVISHVAAV 635
 655 ----NPATNGQLDFMRLSTANTEKSGSAATIKDLNLSQV-----PLTFAGDTPNVTK 706
 636 SETSTDAVNGSQL-----NAVQVQASQPVFTFTGNEGA-VKR 670
 707 KLGEILKVGKKTADDLTKNIGVADSTNLSLTVKLAKT-----477
 671 SLGQSVVISGSSSTAGTSGNLSKVDEAAGRIHLQADSPKPGNVVINGGKISGVTA 730
 748 -LSPLDANVTKL---TASDK--VTVDSGNNTAKL-----QNGDLTFKQNTGA 790
 731 GTEETDAVNFSQLKSI:STAVDQGWTLTASGNGSKVASGGTVDLKNTDENLTISKSGD 790
 791 TPATN-SKTIGVDGLKFTDNNGLALDGTYYITKD--KVGFAKQDGLSKPYLDXKLK 847
 791 DVFNLSKDFKVG-----TSGTTVWVNDGVKVG-----SDVALGTTGLT 831
 848 VGE-VEITNGINAGGKAITG-----LSNTLTDATNATTHGVTOLGIYDSTDKTFAASIG 901
 832 ITDGPVAVTASGIDAGSKVISHVAAVASETSTDAVNG-----SOLNAVQVQASQPVFTT 886
 902 DVLNAGFNLKXNGDAKPFVSTYDVTDFINGNATTAKVTDGKASKAVADVNVVDGTTIHLT 961
 887 ---NEGAVKRESLQGSV-----VISGESSITAG-TYSGNLSKVDEAAG--RIHLQ 930
 962 GADGNKNOIGVKTITLTAKDKKAI:FNFSVNSGDDKALINAKOITADNMLTLAGIRN-T 1020
 931 LADSPK--FG-----NVVING-----GKISGVT 952
 1021 KGTADT-ALQTFQVKKVXKNGDDNDADTITVGDAKNTNOVNTLKLKNGLIDICNTKDG 1079
 953 AGTEETDAVNFSQLKSI-----STAVDQGWTLTASGANGSKVASG--G 993
 1080 TTVTFGINTQSLKAGNNTLNNGLSI:KNTAGNEQIOVGADG--VKFAKVNNGVVGAGID 1137
 994 TV-----DLKNTDGNLTISKSGSDSDVFNLSKDFKVDGMTS 1030
 1138 GTTRITRDEIGFA-----GTNGSLDKSPHLSKDGINAGGKIKTINOSGEIAQNSDAVT 1192
 1031 GTTVWVNDGVKVGSDVALGTTGLTITANGPAVTASGIDAGSKVISHVAAVASETSTDAV 1090

1193 GGIYDLKTELENKISSTAKTAQNSLHFEFSVADEQGNFTVSNPYSSYDTSKTSIDVITFA 1252
 1091 GSOLNAVQVQASQPVFTG-----NEGAVKRESLQGSVVIS-----GESSTACTYS 1135
 1253 GENGITFTKVNKVVVRVIGIDQTKGLTTPKL--TVGNNGKGIVIDSQNGQNTITGLSTLA 1310
 1136 GGN-LKSVVDAAAGRIHLQLA---DSPKFGNVVINGGK-----ISGVT---1175
 1311 NVTNDKGSVPTTQGGKIIKDEDKTRAASIVDVLGAGNLOQNGEAVDFSTYDITVNF--A 1368
 1176 -----AGTBETDAV---NFSQLKSI:STAVDQGWTLTASGANGSKVASGGTVDLKN 1223
 1369 DGNATTAKVTDYDTSKTSKVYVDVNVND-----DTTIEVKDKKLGVKTTTITSTGTGA 1420
 1224 DGNLTISK-----SGSDSDVFNLSKDFKVDGMTSGTTVWVND--GVKVGSDVALGTTG 1275
 1421 NKFALSNOATGDALVKAASDIVAHLNLTLSGDIOTAKGASQANSAGYVDADGNKVIYDSTD 1480
 1276 LTIANGPAVTASGIDAGSKVISHV-----AAGAVSETSTDAVNGSOLNAVQVQASQ 1326
 1481 NKYYQAKNDGTVDKTEVAKDLVAQAQTPDGTILAQMNKSVINKEQ-----1527
 1327 PVTFTG-NEGAVKRS--LGOSVVISGESSTAGTYSGGNLSKVDEAAGRIHLQADSPK 1383
 1528 ---VNDANKKQGI---NEDNAP---VKLEKKAASDNKTKNAAVTVG-----DL 1566
 1384 GNVVINGGKISGVITAGTEETDAVNFSQLKSI:STAVDQGWTLTASGANGSKVASGGTVDL 1443
 1567 -NAVAQTPFTAGTGTAKKLGELTILK-----GGQTDINKLTNNIGVAVG 1613
 1444 KNTDGNLTISKSGSDSDVFNLSKDFKVEITAG-----NTVNTDGVKVGSDVSLGAMGLFIANGPSV 1503
 1614 TDGF-----TVKLAKDLNLSVNAAGTKIDDKGVSFVDSGQAKANTPVLISANGLDLGGK 1669
 1504 HSSTSSVKTLANGESVVNRVTVNGDGVNIDD--VVVNDLGLSIVGASLTLGSLINAGSH 1561
 1670 VISNVGKTXTDAAVQQLNEVENLJLGL--NAGDNADGNQV---NIADIKK-DPN---1721
 1562 KITNVITAGTEDTDAVNFSQLKSVSEAVDKGWTLTASGANGSKVVGSGTVDLKNTDGNLAI 1621
 1722 SGSSNRTVI-----KAGTVLGGKGNNDTEKATGGVQVGVQDKDGNANG-----DLSNV 1770
 1622 SKSGSDSDVFNLSKDFKVDVETAG-----NTVNTDGVKVGSDVSLGAMGLFIANGPSV 1676
 1771 WVKTKDGSKKALLATYNAAGTNYLTNNPA---EADIRINEQGRPHFVNDGNOEPVVQ 1827
 1677 TASGFNAGDK---VISHVAVGMADTDVAVNSQLKQAVQSVTVKATRYYSTNDGG---TQ 1729
 1828 GRNGIDSSASGHSVAIGFQAKADCEAAVAIGROTQA-GNOSIAIGDMAQATGDSOIAIG 1886
 1730 GGNVDGIGATSKAIAAGVGTQASGEGAAAVGSGAAASGKSTAIGRNAIASADGSVALG 1789
 1887 TG-----NVVAKHSK-----AIGDPSTVKADNSYSVGNQNTQFDTAT-----1923
 1790 DGAKDGGGRGABSYTKYSGVQNVNTVGVSVGDAAGKETRSISNVADEKAMDVNLRLD 1849
 1924 -----QT-----DVFGV--GNNTVTES---NSVALGNSAISAGTHAGTQ 1959
 1850 AVAKSNLQTTDDMRHEINNIEDVFKITGDSASSVKGMGVNAMAIGTNAAVS-GTESVAL 1908
 1960 AKKSDGTAGTTTTAGATGVKGFAGQATVAVSVGASGAERRIQNVAAGEVATSOTDAVN 2019
 1909 GKNTVNSADNAVAIG-NGSVADRA-----NSVSVGSGGSEQRVTNVAAG---TABTDVN 1959
 2020 GSOLYKATQGLANATNELDRIHQENKANAGISAMAWASMPQAIIPCRSMVTGCIATH 2079
 1960 VSOL---NQGLITAKQYTDGMVGNLRRRETSGVAAAIAITANLPQAVOGRGMTSVGVSSY 2016
 2080 NGQAVAVGLSKLSNGQWVFKINGSADTQGHVGAAGVAGFHF 2122
 2017 QGOSAIAGVSAVSESHVFKFSANTRSHVGVGAGVGYCW 2059

RESULT 2

A86036
probable adhesin Z5029 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A86036
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86036
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1538 <STO>
A;Cross-references: GB:AE0051174; NID:gl2518349; PIDN:AAG58749.1; GSPDB:GN00145; UWGP:Z5029
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z5029

Query Match 9.2%; Score 986; DB 2; Length 1588;
Best Local Similarity 22.9%; Pred.No.5.3e-31;
Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;

QY 1 MNHYKVIENKANGTFMAVEAYAKSHSTGGSCATCGVGSVRTLSFARIAALAVLIGAT 60
DB 1 MNKIFKVPNATGNTVVTSETAKSGKKSGRSK-----LLISIAVAGM 45
61 LNSAYAGIGISEADGGK-----GANARGDKSIAIGDIAQALGQSIAIGDNKIVHNSNN 117
DB 46 L--SSFGALANAGNDNGQGVYDYGSGSAGDCGWAIGKGA-----NTFMNTSGSS 93
QY 118 NTG--AKASCNESIAIGDVLASGHASIAIGSDDIYKKETVQOISELLPIIRGOKALN 175
DB 94 TAVGYDAIAEGQSSAIGSKXTHAIGGSMAFG-----VSAISE-----131
QY 176 DIYQLADTNLQYRRTHAQCHASTAVGAMSYAKGHFSNAPGTRATAGTYSIAVGLTATA 235
DB 132 -----GDRSIALGASSYSIGQYSMALGRYSKALGKLSIAWGSSKA 172
QY 236 KAASSIAVGSNAQAIGAATAVGGSTQVNLNRGIALGFGSQVLQKDNVNAANVRAYAPD 295
DB 173 EGANAIALGNATKATEIMSLGADUTANASKAYSMAILG--ASSVASEENAI--AIGAETEAEE 230
QY 296 DNQPIDNRYKATPKNG-----ATDVFSIGNSGNDSTIRKLIINVGAGSADTDVNVVA 347
DB 231 NATAIIGNAKAKGTNSMAMFGSLADKVNITIALNGSQALADNAIAIGQ-----NKA 283
QY 348 QLKAEVRLANRQITFFKGDSDNNRVEKGLKTLTITGGAOTSAITDHNIGVVGQDGLKVQ 407
DB 284 DGVDAIALGN-----GSQS-----RGLN--TIALGTASNATGDKSLAGSNS-- 324
QY 408 LAETLSLKMVTENITANEKVVTVGKTRITTDKIGFTNDMNGIDESKPYLDKDTG--IHA 465
DB 325 -ANGINSVALGADSIADLNDTVSGNSSL---KRKIWNKNGAIKSDSY--DAINGSQLVA 379
QY 466 GGQKIYKLTITAG--VVDLDAATYQOLKKVKNQTAESALQTFTVKKVDKNGDAN---DSKI 519
DB 380 ISDSVAKRGLGGGAADVDDGTV-----TAPTNYLNKSGSKNNVGAALAVLDENT 427
QY 520 I-----TVGKNNKPDGTQVNTLKLKXGNGVDVTTETNGTFTFGLNQNNGLTVGNSILNNDG 575
DB 428 LQMDQTKGKYSAAGHTSSPTASV-----ITDVAAGTIS-----460
QY 576 LSVKNVNSNKQIQVGADGITFTDISNPKGAGIENTTRITRDGIGFANNTGSLDANKPRL 635
DB 461 ASSKDAVNGSQLKATNDDE-----ANTANIATNTSNIATNATNTNI 507
QY 636 TPTGINAGGKELINVQSAINPATNGGOLDPMNRLSTANTEKSSAATIKDLYNLSQVPLT 695
DB 508 T-----NLDTSVGLDQ--ADALLWNETKKAFAA-----534

Db	1312	-----KVTDAQVSQNTQSITN-----LDNRVTNLDSEVT--NIENG	1345	Db	173	EGANAIALGNATKATEIMSIATGDTANASKAYSALG-ASSVASEENAI-AIGAETEAAE	230
Qy	1757	VDKXGNANGDLNVVVTQKDGSKKALLATYNAAQTNYLTNNPABAIIDRINEQIRFFH	1816	Qy	296	DNQPIDNRYKATFNG-----ATDVFSIGNSGHDSIRRKLIINVAGSADTDVAVVA	347
Db	1346	I-----GDIV-----TTGSTKYFKTN-----	1361	Db	231	NATAIGNAKAKGTNSMAMGFGSLADKVNTIALGNGSQALADNAIAIGQG-----NKA	283
Qy	1817	VNDGNQEPVQGRNGIDSSAGKHSVAIFGQAKADGEAAVAIGRQTQAGNQSIAIGNAQ	1876	Qy	348	QLEAEAVELANRQITFKGDDSNRVEKGLGKLTLTITGGAQTSALTGDHNIQVGVQNGDGLKVQ	407
Db	1362	-----TDGVDPASAQKDSVAI-----GSGSIAAADN--	1387	Db	284	DGVDATIALGN-----GSGS-----RGLN-----TIALGTASNATGDKSLAGSNSS--	324
Qy	1877	ATGDQSIAIGTGNVAGKHSIGAIDPSTVKADNSYSVGNNNQFTDATQTQDVFVGNNTV	1936	Qy	408	LAETLTSLKMMVTENLTANEKVTVGKTRLTDTDKIGFTNDMNGIDESKPYLDKDTG--IHA	465
Db	1388	-----SVALGTGSV-----ATEENTISVGSSTNQ	1411	Db	325	-ANGINSVALGADSIADLDNTVSVGNSSL--KrkIVNVKNGAIKDSY-DAINGSOLYA	379
Qy	1937	TESNSVALGNSAISAGTHAGTQAK-KSDGTAGTCTTTAGATGTVKGFAGQTAVCASVGA	1995	Qy	466	GGQKITKLTAG--VVDODAATYQOLKKVQNTAESALQTFVTKVVDKNGNDAN-----DSKI	519
Db	1412	RRITNVAAGRNA-----TDVNVAAQLKSSBAGGVRYDTKADGSD--YSNITLG-----GG	1460	Db	380	ISDSVAKRLGGGAADVDDGTV-----TAPYLNKNGSKNVGAALAVLDENT	427
Qy	1996	SGBERRITONVAAGEVSATSTDAVNGSOLYKATQCIANATN-----ELDRHIONENKANAG	2051	Qy	520	I-----TVGKNNKPDGTQVNTLKLKGENGVDTTETNGTFTFGLNQNNGLTVGNSTLNDNG	575
Db	1461	NGGTIRLSNVAG--VNNDVNVYAQLKQSVQETKQYTDQRMVEMDNKLSKTESKLSGG	1517	Db	428	LQWDQTKGKYSAHGTSSPTASV-----ITDVADGTIS-----	460
Qy	2052	ISSAMAMASPMQAYIPGRSMVTGGIATHNGGAVAGLSKLSDRGQWVFKINGSADTQGH	2111	Qy	576	LSVKNTNSNKQIQVADGIIFTDISNSKPGAGIENTTIRTRDGI GFANNTGSLDANKPRL	635
Db	1518	IASAMAMTGLPQAVTPGASMASIGGTVNGESAVALGVSVMVANGRWYKLGQSTNSQGE	1577	Db	461	ASSKDAVNGSQLKATNDVE-----ANTANIATNTSNIATNTANIATNTTNI	507
Qy	2112	VGAAVGAGFHF 2122		Qy	636	TPGTINAGGKELTNVQSAINPATNGGQOLDPMFNRLSTANTEKSGSAAITIKOLYNLSQVPLT	695
Db	1578	YSAALGAGIQW 1588		Db	508	T-----NLTDVSGDLQ--ADALLMNETKKAFSA--	534
RESULT 3				Qy	696	FAGDTGPNVTKLGEILKVGGKTTADDLTKNNIGVVADSTDSNLSLTVKLAKTLSDLDVN	755
H91188				Db	535	-----HGQDTTSK-----ITNVKDA-----DLT-----ADSTD-AVNGSQLKTTDAVATN	574
probable adhesin ECs4480 [similarity] - Escherichia coli (strain O157:H7, substrain R1MD				Qy	756	TKLTATASDKVTVDSGNNVAKLQNGDLTFSKONTGATPATNSKITGVGLKFTDNGI--A	813
C;Species: Escherichia coli				Db	575	TTNI-----ANNTSNIATNTNISNLTEVT-----NLGEDALKWDKONGVFTA	618
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001				Qy	814	LDGTYITKDKVFAKQDGSIDK-SKPYLDKDKLVGEVEITTINGINAGGKAITGLSNTL	872
R;Accession: H91188				Db	619	AHGTE--TTSKITNVK-DGDLTTGSTDAVNGSQLKTTNDVATNTIATNT--TNISLIT	674
R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.				Qy	873	TDATN-----ATTGHVTQLGIVDSTDKRAASICDVINAG--FNLKNG	914
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.				Db	675	ETVNLGEDALKWDKONGVFTAAGNNTASKITNILDGTVTATSSDAINGSQLYDLSN--	733
DNA Res. 8, 11-22, 2001				Qy	915	DAKDFVSTYDTPFINGNATTAKVTVYDGKASKVAYVNVVDGTTTTLTGADGNKKNQIGVKT	974
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno				Db	734	-----IATY-----FGGNAS-----VNTDGVFTGPTYKIGETNYVNV--	765
A;Reference number: A99629; MUID:21156231; PMID:11258796				Qy	975	TTLTKTDAKD--KAIN--FSVNSGDDKALINAKDIADNLTLAGIRNTKGTADTALQT	1030
A;Accession: H91188				Db	766	-----GDALAAINSFSFSLGD--ALL-----WDATAGKFSAKHGT-----	799
A;Status: preliminary				Qy	1031	FQVKVXKENG-----DDNDADTITVGKDAKTNOVNTLKLKGN-----GLDIQTNKD	1078
A;Molecule type: DNA				Db	800	-----NGDASVITDVADGEI-----SDSSDAVNGSQLHGVSYYVVDALGGGAENVAD	847
A;Residues: 1-1588 <HAY>				Qy	1079	GTVTFGINTQSGLKAGNNTLINNG--LSIKNTAGNEQIQVGADGVKFAKVNNGVVGAGI	1136
A;Cross-references: GB:BA000007; PIDN:BAR37903.1; PID:g13363955; GSPDB:GN00154				Db	848	GTITAPTPTTIA-----NADYDNVGDALNAIDTTLDDALLMDAD-----	885
A;Experimental source: strain O157:H7, substrain R1MD 0509952				Qy	1137	DGTTIRTRDEIGFAGTNGSLDKSKPHLSKDGINAGGKKTNIOSGEIAQNSNDVATGSKI	1196
C;Genetics:				Db	886	-----AGENGAF--SAH-----GKQKTSVITNVANGALSAASSDAINGSOL	926
A;Gene: ECs4480				Qy	1197	YDLKTELENKISSAKTAQN--SLHFEFVADEQNNFTVSNPYSSYDTSKTSBIVITPAG	1253
Query Match				Db	927	YTTNKYIADALGGDAENVADGTITAPTITIANAEVNN-----VGDAIDALD	972
Best Local Similarity				Qy	1254	ENGITTKVNGVVRVGIDQTKGLTTPKLTGVNNNGKGIIVDSQNGQNTITGLSNTLANVT	1313
Matches 510; Conservative 273; Mismatches 696; Indels 752; Gaps 95;							
Qy	1	MNHYYKVFENKATTFMAVEAYEAKSHSTGGSCATGQVGSVRTLSFARIAALAVLIVIGAT	60	Db	1	MNKIFKVINPATGNYVTSETAKSRKKSGRSK-----LLISALVAGM	45
Db	61	LNGSAYAGIGISEADGGKG---GANARGDKSIAIGDIAQALGQSIAIGDNKIYHNSNN	117	Qy	46	L--SSFGALANAGNDNGQVDVYSGSGAGDGWVAIGKGA-----NTFMWTSGSS	93
Db	118	ANIG--AKASGNESIAIGDVIASGHASIAIGSDDLYLKRETVOQISELLPIIRGOXALN	175	Qy	94	TAVGYDAIAEGQYSAIGSKTHATGGASMAFG-----VSAISE-----	131
Db	176	DIYQLADTNLQYRTRTHAQGHASTAVGAMS YAKGHFNAPGTRATAGCTYSLAVGLTATA	235	Qy	132	-----GDRSIALGASSYSLQYNSWALGRYSKALGKLSIANGDSKSA	172
Db	236	KAASSIAVGSNAQAIQFAATAVGGSTQVNLNRGIALGFGSQVLQKDNVDNAANVRAYAPD	295	Qy			

Db 973 DNALLW-----DETA-----NGGAGAYNASHDKKASI-----ITNVA 1004
QY 1314 NDKGSVRTTEQGIKIKDEKTRAAISIVDLVSAGFNLOGNEAVDFVSTYTVNPAQGNAT 1373
Db 1005 N--GSI-----SEDS-----DAVNGSQNLAT 1024
QY 1374 TAKVTYDDTSTKSKVVDV--NVDDTTIE-----VKDKLGVKTTTLTSTGTGANKF 1423
Db 1025 NWMI-----EONTQINQLAGNTDATYIOENGAGINVRINDGCLAFDASAOQVGNAT 1079
QY 1424 ALSNQATGDALV-----KASDVAHLNLTSGDIQT-----AKGASQANSSAGYVDADGNKVI 1475
Db 1080 GYNSVAKGSSVAIGQSYSDVTGTIALGSSSVSSVRIAGSRDTS-----ITENGVIIG 1134
QY 1476 YDSTDNKYQAKNDGTVDTKEVAKOKLVAAQPTDGLAQMNVKSVINKEQVNDANKQ 1335
Db 1135 YDITDELGLALSIG-----DDGKYRQ 1156
QY 1536 GINEDNAFVKGLEKASDNKTKNAAVTVGDL-----NAVAQTPPLTF--AGDTGTTAKKIGE 1589
Db 1157 IIN-----VADGSEADAVTVRQLONAIGAVATPTKYPHANSTEEDSLAVGT 1204
QY 1590 TLTIKGGQDTNKLNNIGVAGT-----DGTVVKLAKDLTNLSNVNAGTKIDDKG 1642
Db 1205 DSLAMGAKTIVN--GDKGIGIGYCAVYDANALNGIAIGSNAQVIHVSIAIGNSTTTRG 1262
QY 1643 VSF-----VDSGQAKANTFVLSANGLDLGGKVISNVGKGTDXDAAVQOOLNEVRNLL 1696
Db 1263 AQNTYTYNMDAPONSVEPFSVSAD-----GQROIINVAAGSADTDAVNVGQL----- 1311
QY 1697 GLGNAGNDAGNQVNIADIKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLAGTVGVG 1756
Db 1312 -----KVTDQAQVSONTQSIIN-----LDNRVNLDSRVT--NIENG 1345
QY 1757 VKDGNANGDLSNVWVKTKQDGSKKALLATYNAAGQNTYNTNPAEALDRINEQGRFFH 1816
Db 1346 I-----GDIV-----TTGSTKYFKTN----- 1361
QY 1817 VNDGNQEPVVQGRNGIDSSAGKHSVAIGFQAKADGEAAVAIGRTQAGNQSTAIQDNAQ 1876
Db 1362 -----TDGVDAQAQKDSVAI-----GSGSIAAADN-- 1387
QY 1877 ATGDQSIAGITGNVAVAGKHSIGAIDPSTVKADNSYSVGNNOFTDATQDVFVGNNTIV 1936
Db 1388 -----SVALGTGSV-----ATEENTISVGSSTNQ 1411
QY 1937 TRSNSVALGSNSAISAGTHAGTQAK--KSDGTAGTTTTAGATGTVKFGAGTAVGAVSVGA 1995
Db 1412 RRITNVAAGKNA-----TDAVNVAQLKSSAGGVRYDTKADGSDID--YSNITLIG--GG 1460
QY 1996 SGAERRIQNVAAAGEVSATSTDAVNGSOLYKATQGIANATN-----ELDHRHONENKANAG 2051
Db 1461 NGTTFRISNVAG-----VNNNDVVNVAQLKQSVQETQYTDQRMVEMDNKLSKTESKLSGG 1517
QY 2052 ISSAMAMSPQAYIPERSMVTGGIATHNGQGVAVGLSKLSDNGOWFKINGSADTQGH 2111
Db 1518 IASAMAMTGLPQAYTFGASMASIGGGTYNGESAVLGVSMVSANGRWVYKLOQSTNSQGE 1577
QY 2112 VGAACVAGFHF 2122
Db 1578 YSAALGAGIQW 1588

RESULT 4
A82615
surface protein XF1991 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82615
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82615; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82615
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1190 <SIM>
A:Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN003
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; F. Briones, M.R.S.; Bueno, M.F.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, F. as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; Laig chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasah A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; ;
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1981

Query Match 6.8%; Score 728; DB 2; Length 1190;
Best Local Similarity 23.3%; Pred. No. 4.4e-21;
Matches 354; Conservative 191; Mismatches 516; Indels 458; Gaps 59;

QY 707 KLGEILKVGKGTADDLTKNNITGVVADSTDNLSLTVKLAKTLSDLDVAVNTKLTITASDKVT 766
Db 27 KIGOSIQ-HAVMTSAASSKK---GTQPRRSNNAMTAKRSAKLHRRQLRHIVLLTVLAAS 82
QY 767 VDSGNTAKLO-NGDLT-----FSKQNTGATPATNSKT-----IGVDGLKFTDNGIA 813
Db 83 GYTGVAAQVYVNSDSTENCVEILGDSQSQTFSHTASNDKCKPDFQTQTEYSLEYDYENLV 142
QY 814 LDQTTVITKDKVGFAGQDGLSKSPYLDKDKLKVCEVEITTINGINAGGKAITGLSNLT 873
Db 143 LGSLLVNEGKGLVDISGA-----TYSMLGSI-ATMNG-SAGIDSIAGSQGS 191
QY 874 DATNATTHVHTQIGIVDSDDTKTRAASIG-----DVLNAGFNKNGNDKADKFVSTYDT 925
Db 192 KTDGNTSGATVAQGL-----RSIAIGTTARSQSDAISIGTGASTGNFAIANGAL 244
QY 926 VDFINGNATKATVYDCKASKVAYDVNVDGTTIHLTGADGNKQKQIGVKTTLTKTDAKGD 985
Db 245 TSIANGIALGASSVTTTRGGVALGQGSAAATASGITGYD-----PVTKSTS--- 290
QY 986 KAINFSVNSGDDKALLINAKDIADNLNLTAGEIRNTKGTGTA---DT-ALQTFQVKKVKENG 1040
Db 291 -----TLTSMWRSTLGAVSIG-NITSTSTQRTGLAGRSDDTDAVNVQKLL--- 340
QY 1041 DDNDADTTIVGDKAKTNQVNTLKLKGNGLDIQTNKGGTTFGINTQSLKAGN-NTTL 1099
Db 341 -----AESVGGWNLTFASGANSNVALGESVDLK-NSDCNLLITKTTDSNDVTFNLATAL 394
QY 1100 NNGLSIKNTAGNEQIQVAGDGVKFAKVNNGVVGAGIDGTTITRDEIGFAGTNGSLDKS 1159
Db 395 KYDSLTTGNTA-----MTTDTGVTGK-----RVTLDDSTGLVIAEG----- 429
QY 1160 KPHLSKDGINAGKKITNIQSGEIAQNSNDVATGGKIYDLKTELENKISSTAKTQNSLH 1219
Db 430 -PSVTSSGINAAGQKLMNVGTGTA---DTDAVNFQGL-----QAVSDTA----- 469
QY 1220 EFSVADEQGNFTSVNPFYSSYDTSKTSVITFAGENGITTKVNGVVRVIGIDQTKGLTTP 1279
Db 470 ----- 469
QY 1280 KLTGVNNKGGIVDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGIKIKDEKTRAASI 1339
Db 470 -----SKGNLLASGTSNNVAPGASVDLKN-----TDGNLLI-----TKAIGI 508
QY 1340 VDLSAGFNLOGNEAVDFVSTYTVNPAQGNATKATVYDDTSTKTSKVVDVNVDDTTI 1399

Db 509 NDVT---FNLATALEA-DSLTTGNTAMTTDGVTVGSNVT---LGSTGLVITD----- 553
Qy 1400 EVKDKLVKVTTLTSTGTGANKFALSNQATGDA---LVKASDIVAHNTLS-CDIQFATK 1455
Db 554 -----GpsvTSSGSIAGNOKITNVAAGTADTDAVNFPSQLQAVSSSTASKGNWLLAS 603
Qy 1456 GASQAN---SSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKVAKKLVIA--CAQT 1509
Db 604 GANSNVVPGESVDLKNSDGNLLITKTIDS-----NDVTFNLATALKVSLTTGNTAMT 657
Qy 1510 PDGTLAQMNKVSINKEQVNDANK--KQGINEDNAFVKLEKAASDNKTNAVAATVGDNLN 1567
Db 658 TDGVTVGSNVTLGSTGLVITDGPVSTSGI-----SAGNOKITNVAAGTADT 705
Qy 1568 AVAGTPLTFAGDTGTTAKKL-----GEFLITKGGQTDNKLTDNNIGV--VAGTGDFTVKL 1621
Db 706 AVNFSQLQAVSSTASKGNWLLASGANSNVAPGESVDLKNKTGDNIVISKESGSDNVLN 765
Qy 1622 AKDLTNLSNVNAGGTTKIDDKGVSF-----VDSSGOAKANTPVLKANGLDLGGKVISNVGK 1676
Db 766 SSSL-KDKLTVGTVMTNGVTVGSVTLGSMGLVITDGPVSTSSGINAGSOKITNVA 824
Qy 1677 GTKDDTAANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRTVVIKAGTV 1736
Db 825 GTADTDAVNLSQLN-----GGNYNGD-----TA 840
Qy 1737 LGKGKNDTEKLTAT--GGVQGVVDKNGANGDLSNVVVKTKQDKGSKALLATYNAAGQTN 1794
Db 841 MAGSGAKSVHYSTYDGGTQ-----GGNYNGD----- 867
Qy 1795 YLTNNPABAI DRINEQGRFFHVNDGNQEPVVGQGRNGIDSSAGKHSVAIFQAKADGEA 1854
Db 868 -----CATGTRSIANGVGTLASAEG 887
Qy 1855 AVAIGRQTA--GNOSIAGDQAQATGDSIAIGTG-----NVVAKHSG----- 1897
Db 888 ATAVSGAAAGKGSSTAIGRNAVASADGSVALGDKAGDARGAESYTKYSGLQNNVTGT 947
Qy 1898 -AIGDPSIVKADNSVSGNNQFTDAT---QTDVFG-----VGNNI-----TVTESN 1940
Db 948 VSGDASKGERTVTSNADAKEDAVNLRLDRVAQDANRYVDNKIESLSEGGTFFVKVN 1007
Qy 1941 SVALGNSA--ISAGTHA-----GTOAKKSDGTA--GTTTTAGATCTVKGFAGQAVG--- 1989
Db 1008 SL-----NNSATPIAGVDATAIGVGATASGASDIANGKASASADNAV-----AIGHNS 1057
Qy 1990 -----AVSUGASGAERRIQNVAGEVSATSTDVANGSOLYKATQGIANATNELDRIHQ 2043
Db 1058 VADRANTVSVGSAGSERQVTVNVAAG---TADTDAVNVSQL---NQGLITAKQYTDGVVGS 1111
Qy 2044 NENKANAGISSAMAMASMPQAYIPGRSMVTGGIATHNGQAVAVGLSKLSNGOWVFKIN 2103
Db 1112 LRRDTGGVAAATATANIPQAYIPGRGMTSVGVSSYRQGSATVAGVSSVSGRWVFKFS 1171
Qy 2104 GSADTQGHVGAAGVAGFHF 2122
Db 1172 GSANTRSQVIGAGVGYQM 1190

RESULT 5
AC0976
probable autotransporter sapB [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0976
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, J.; Parkhill, J.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Farry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AC0976
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1107 <PAR>
A;Cross-references: GB:ALU513382; PIDN:CAD03303.1; PID:g16504923; GSPDB:GN00176
C;Genetics:
A;Gene: sapB

Query Match 6.38; Score 670; DB 2; Length 1107;
Best Local Similarity 23.74; Pred. No. 7.3e-19;
Matches 304; Conservative 193; Mismatches 413; Indels 372; Gaps 59;

Qy 993 NSGDDKALI--NAKDIADN---LNTLAGEIRNTKGTAT---DTALQTFQVKKVKGNGDDN 1044
Db 46 NAGNDTGSINQNTTDTATNTTSINNLSNVTTLTDDALLWDAASGTFASR--NGSASK 102
Qy 1045 DADTI--TVGDKAKTNQVN-----TULKKGKGLDITQKDGTVGTGINTQSGKAGNN 1096
Db 103 ITNLAAGTAAADS--TDAVNGSQLFDNTEKVDQNTADITTN-----TNSINQNTTDTATNT 156
Qy 1097 TTLNNGLSIKNTAGNEQIOVGADGVKFAKVNNGVWGAGIDGTTTRIDRDEIGFAGTNGSL 1156
Db 157 TSINN-----LSNSV-----TTLTDDALLWDAASGTF 183
Qy 1157 DKSPHLKSDGINAGGKKITNIOGGEIAQNSNDVATGKIIYDLKTELENKISSAKTAQN 1216
Db 184 SASR-----NGSASKITNLAAGTAAADSTDAVNGSQLFDNTEKVDQNTADITTNTN 234
Qy 1217 SLHEFSVADQGNFTVSNPYSSYDTSKTSVI-----TF-AGENGITTKV-NKGVVR 1267
Db 235 SINQNTT--DIATNTTSINNLSNVTTLTDDALLWDAASGTFASRNGSASKITNLAAGT 292
Qy 1268 VGIDOTKGLTTPKLTGCVNNGKGIVIDSON--GQNT--ITGLSNTLANVNDKGSVRTTEQ 1324
Db 293 LAADSTDAV-----NGSOLYETNOKVDQNTSAIDINTSITNLSSD----- 333
Qy 1325 GKIIKDEKTRAAASIVDVLGAGFNLOGNEAVDFVSTYDVTNFDAGNATTAKVTVDDTSK 1384
Db 334 -----NLSWN-----ETTNSPSAGHSSTTKNTKTNVAAGE 363
Qy 1385 TSKVYVYDVNDTTIEVKDKKLVKVTTLTSTGTGANKFALSNOATGDALVKASDIVAHL 1444
Db 364 LSEBSTD--VNGSOLFETNEKVDQNTDIAANTTN-----ITQNSTAIENL 408
Qy 1445 NTLSGDITAGASQANSSAGYVDADGNKVIYDSTDNKYQAKNDGTVDKTKVAKOKLV 1504
Db 409 NTSVSDINT-----SITGLTD--NALLWDE--DTCAFSAHGGSTSKITNVA----- 451
Qy 1505 AQAQTPDGTLAQMNKVSINKEQVNDANKKQGINEDNAFVKLEKAASDNKTNAVAATVVG 1564
Db 452 AGALSSEDST-----DAVNGSOLYETNOK-----VDQNTS-----AIA 483
Qy 1565 DLNVAQTPLTTFAGDTGTTAKKLGETLITKGGQTDNKLTDNNIGVWAGTGDFTVKLAKD 1624
Db 484 DIN-----TSITNLGTDALSDWDDERGAFAFSAGTSGTNTKITN-----VAAG-----EIASD 529
Qy 1625 LTNLSNVNAGGTTKIDDKGV---SFDVSSGQAKANT--PVLKANGLDLGGKVISNVGKGT 1679
Db 530 ST--DAIN--GSOLYETNMLISQYNESISQLAGDTSETYITENG--TGKVIKIRNDNGLE 583
Qy 1680 DTD-----ANVOQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSRTVTKA 1733
Db 584 QDAYATGATGATVGYDAVASGAGCALGQNSSSSSIEGIA-----LGSSTSRRAITTG 638
Qy 1734 -----GTVLGGKNNDEKLTATGQVGVVDKNGANGDLSNVVVKTKQDKGSKALL 1784
Db 639 IRETATSDDGVVI---GNTTIDRELLGALSUGT--DGESYQITNV-----ADGSE----- 684
Qy 1785 ATYNAAGTNTLTNNPABAI DRINEQGRFFHVNDGNQEPVVGQGRNGIDSSAGKHSVAI 1844
Db 685 -----AQDAVTVQLQNAIGAIVTTPTPKYHANSTEB-----DSLAVGTDLSLAM 728

QY 244 GSNAQAIQPAATAVGSGTQVNLNRGIALGFGSOVLQKQNDVNAANVRAVAPDDNQPIDNR 303
 Db 263 GD-----IG-----NIGSLDITNVLGSAF-FNSTILKATNINLKHNTSTNLNDN----- 307
 QY 304 YKATFKNGATVFSIGNSGNDISIRKIIN--VGAGSADTDVAVQAQLKEAVELANROIT 361
 Db 308 -----IIVIGNIKGN--XDLNFKVHGHTNLDNEMIIIPAPQ-----THGTEN 349
 QY 362 FKDDSNV-----RVEKGLGKLTITGGAQ-----TSA 399
 Db 350 FKGNATLNGNINNLNFKSGHGKTLNQGNTKVDNLVPADSVLDSGTISVNGLLDTC 409
 QY 390 LTDHNIQVWNGDGLVQLAETL-----TSLKMTTENTLANTANEKVTVGKTLRTLDKIG 442
 Db 410 VTFNNSV--NGTLTINAKNTISAKLNLAKAKIQINANLWNH-----PSAG 456
 QY 443 FTMNMGIDSKPYLDRDTGIHAGGOKITKLTAQVVDVDDAATYQOLKKVNTQTAESALQTF 502
 Db 457 DISDIRIADNTIYIDAKN---GNVLLNNAKIIPEGADS--MLALIN-TGVADRTF 509
 QY 503 TV-KKVDKNGDAND-----SKITVGNKPKDGT-QVNTLKLK----- 539
 Db 510 TIYNLNQSGNDEYIGVIEAKKVTITIANQSGPYTIGQDNTHLRELIVGAGDIIIDD 569
 QY 540 -----GENGVDDVTETNGTVTFGLNQNNGLTVGN 568
 Db 570 TIFTKLLINSTQITFNRTLDLQAGNIAFGHGLVLVNGVTGSIITSENNOGILITNS 629
 QY 569 STLNDGLSVKNTNSKQIQVGADGIT-----FTDISNSKPGA-----GI----- 608
 Db 630 G--NITGVIGTNELGLKLVNIGADPVCNSANVPASVALTNPSSVLIADGVTLTGCVTH 687
 QY 609 ENTRITRIGIPANTGSLDANKPRLTPPTGINAGGKELTNVQSAINPATNGQQLDFMNR 668
 Db 688 NNTKGVLSLGTG--SNITGQITNSAALKINIGAGA---SNIDSNIYAG----- 732
 QY 669 LSPANTEKSSAATIKDLYNLSOVLPTFAGDTGPNVTKLGEILKVKGGKTA--DDLTKN 727
 Db 733 -STVLTDQITSELTLNDVVVNSNIIITTAGNSKGLIFTGNGGITGNIANGAALQEVFN 791
 QY 728 --NIGWADSTNSLTVKLAK-----TSLDLDVANTKLTATASDKVT--VDSGNNTAKL 776
 Db 792 GTTNIGGTANSQNTVAHSAANVITGLTGLKLYKXDTGIIAHGGLVGDIDFNKAGKF 851
 QY 777 QNGD-----LTF-----SKQNTGATPATNSKTIG----- 800
 Db 852 ILGDGAMIDGSLCNGVAGTLDFIGDGNVTQNIAGDANANSISTINIQDNTKNVTIAND 911
 QY 801 --VDGLKFTDNNGIALDGTIYITKDKVQPAQDGLSKPKYLDKDKLVGEVEITNGI 858
 Db 912 IFVDNTHFTNGGLIQLGNN--LTHNIDFGANGGTLFN-----GNNTYLNAL 958
 QY 859 NAGGKAITGLSNTLTD--ATNATTG-----HVTQLGI-----VDSDTKRAASIGDVL 904
 Db 959 IVNGQ--NGLNAFTNLKASDDTIGTVKIINIQTGTPQNTFIQVANKNLTIVSSVSSI 1016
 QY 905 NAGFNLKNGDAXD-----FVSTYDVTDFING-NATTAKVTYDGKASVAYDVNVGDTTH 959
 Db 1017 NF-----GDANSQILSAPVDQITKFINNLNETGGI-----ITLDSN--GNLNT 1059
 QY 960 LTCADGNKNOIGVKTTLTKTDAGDKAI-----NFSVNSG--DDKALINAKDI 1006
 Db 1060 ISGNNGIK--LGSKNELSLNIGKVTVNDLIDIQNIHQINNGALFDDQSLTSKAKIK 1117
 QY 1007 ADNLNTLAGIRNTKGTADTALQTFQVKYKENGDDNDNA-----DTITVGKDAKTNOV 1060
 Db 1118 NINIGTVAGG-----ATYTLDAINDFNLTSQWFKHQDSILLEKNSNTND 1165
 QY 1061 NTLKKGKNGLDIOTNKDGVTVFTGINTQSLGKAGNNTLNNGLSINKNTAG--NEQIQVG 1118
 Db 1166 HTITL--TSALDPGNNQFGIILITDT-----NKLITIDNNG-NVAYTLTGTAHMLK-- 1213

QY 1119 ADGVKFAKNVGV---VGAGIDGTTTRTD-EIGFAGTNGSLDKSPHLSKDGINA--- 1170
 Db 1214 --QLTFASIDNGAIALKVGINVENTVNLKIDIELNEVNAVNLFNKNTTYTATGNGHVD 1271
 QY 1171 --GGKKITNIQ-----SGEIAQNSDAVTG-----GKIYDLKTELENKISSTAKTAQN 1216
 Db 1272 FQGNAGVINLDDIEDIGSVTSGN--VNGTLNFGNSGKVTGLINNIIVMLQAGAGDVLS 1329
 QY 1217 SLHFSVADBOG---NNFT-VSNPYSSYDTSKT-----SP 1247
 Db 1330 ASGNSITBIQNGNNNLTPAANSHTLTDINKTGGQDLNLVFNINGSVSSIGANAAGVD 1389
 QY 1248 VITFAG-----ENGITTKVKNKGVVRVID--QTKGLTTPKL----- 1281
 Db 1390 IINAGSVNFSNTLKSNGNIVISDGAITQVNNVATDISEKQANNGLKLNHTPINITS 1449
 QY 1282 TVGNNGKGVIDSQNGQNTIT-----GLSNTLANVTNDKGSVR 1320
 Db 1450 TLGNNAIG--TIEVANNDDVTITGTLOAQNIHFSNATQAATLTILGAASQVTNITTAGNNIH 1508
 QY 1321 TTEQGLIIKDEDKTRAASIVDVLS--AGFNLOQNG---EAVDFVSTYDVTNPAADGNA-- 1372
 Db 1509 TLE---VTFDFTGNDGIIGDANNRLKSIELTGNGTVINSPHYSSITANNAAQGNVL 1564
 QY 1373 -TTAKVTYDTSKTSKVVDVNVDDTTI--EVKDKKLGVKTTTLTSTGTGANKFALSNOA 1429
 Db 1565 NIEGGITYDLGSKIKSLANVOISEDTIRGDVYSKVLNIDAGKTINFRDGNMNNPKND 1624
 QY 1430 TGDALV-----KASDIVAHLNLTSGD--TOTAKGASQAN--SSAGYVDAD--GNKVIVDSTD 1480
 Db 1625 IPDALIDLVLPRSLFNFYFTDIKADNLNFADDTATANKDAVWIDAHIDNGGILKFND 1684
 QY 1481 NKY--YQAKNDGTVDKTEKVAQKLVQAQTPDGTGLAQMNVS-----VINKEQ 1527
 Db 1685 NAWLTOEIKNANI---ETASDKF-----LLOKNIKATLIADNANILVLDNVE 1731
 QY 1528 VNDANKKQGINEDNAFVKLEKASDNKTNAATVTGDLNNAVAQTPLTFAAGDTGTTAKL 1587
 Db 1732 VNTNLNVRDIVLDA--NYELKYTGNTVTHGLTTI-----ITY---PDALQKG 1775
 QY 1588 GETLTIKGGQDTNKLTDNNIGVAGTGTGTVKLAKDLNLNLS-----VNAGGTX 1637
 Db 1776 GHILVSGSNVMDSL--DNLI-----IKIKAHSDITNITSDTKHQIVKLETGAIYTP 1826
 QY 1638 IDDKGVSVFVDSQQAANTPVLISANGL---DLGGKVISNVG-----KGTKDTDAANVQ 1688
 Db 1827 VPQTKVIIIDASEBQNKVFWADANGLVLLTDTGGR--DDTGRDDTGRGNTD----- 1878
 QY 1689 LNEVRNLLGLGN-----AGNDNADGNQVNIADIKK--DPN-----SGSSS 1726
 Db 1879 -NGCRDNCVGNLSNNSNEAGSSSDKN-YGITDVVPFDFPSPILDYTKNNYVAGIAN 1936
 QY 1727 ---NRTVIKAGTVLGGKNNDETEKATGVGVQVQDKG---NANG-----DLSNVWVK- 1773
 Db 1937 QLINHVKDFGNTTDAGKLLNDLGFMSPNRVTTDLRLSNRINVLNGLNEGVGLNGIEVEN 1996
 QY 1774 -----TQKD--GSKKALLA---TYNAGOTNYLTNNPAEADIRNEQIRFFHV 1817
 Db 1997 FLTDIAINMDNFTAKSIGRLEBELSDANTVGNLTKNTLNK-----INLKL 2045
 QY 1818 NDNQNPVVQGRNGIDSSAGSKHSVAIFQAQKADGAAVAIGRQTAQNGSIAIGNAQA 1877
 Db 2046 NTNNOAIAGDE--DNIVTIGWMSFYCKIKONSNS--ASGVQSNVTGGGILGFYDNI-- 2100
 QY 1878 TGDQSIAITGNVAVG---KHSGAIGDPTVKADNSYSVG-----NNQFTDA----- 1922
 Db 2101 --DNSIVIGAAVTMAQSKVKKNDKNGDRTKAKSNISYISYGLYNLNTNFFVEAIGVYGR 2158
 QY 1923 -----TQTDVFGNNITVETESVALGNSAISACT 1954
 Db 2159 NKIKNYEKRTITTDQIAIGKFINTFYSYELLGGYNYLISHRT 2201

RESULT 8
F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R:Hayashi, T.; Makino, K.; Kurokawa, M.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033965.1; PID:gl3360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: ECs0542

Query Match 6.0%; Score 645.5; DB 2; Length 5291;
Best Local Similarity 21.9%; Pred. No. 5.3e-17;
Matches 582; Conservative 293; Mismatches 932; Indels 855; Gaps 131;

Qy 35 TQGVSVRTLSFARIALAVLIVGATLN---GSAYAGIGISEADGG-KGGANARGDKSI 89
Db 1136 SGPLTINTVSGDDIINAAEIVVAQTISGQVGTAVAGTIVITIGGNQYNATVQSDLSW 1195

Qy 90 AI---GDIQAALGSQSTAIGDNKIVHNNANIGAKASNESIAIGDVLASGHASIAI 146
Db 1196 SVSPANVLQALNGELTISAS--LTNSANN-----TGTATHDIVIDANLPLGLRVDV 1246

Qy 147 GSDLLYKKEVTVQOISELLPIRGOKALNDIYQLADTWLQVRRTHAQGHASTAV---- 201
Db 1247 AGDDVINSIEHTQAL-----VITGSS-----GLAAGALTIVINSVT 1284

Qy 202 -GAMSYAKGHFSNAPGFRATAE---GTYSVLAVGITATAKAASSIAGVNAQAIGFAATAV 257
Db 1285 YGATVLADGWSVGPVADVNTWPAQTIVIAVSGTNTAGTTTSI---SHPVTVDLAAVAI 1341

Qy 258 -----CGSTQVNLNRGIALGFGSQ-----VLQKN-----DVNA 286
Db 1342 TINTLSTDDVINAAEKSGDQLSGLTSGVEAGQITVIFGKSYTTTVAADNTWGLTIPA 1401

Qy 287 ANRVAPYDDNQPIDNRKYATFKNGATDV-FSIGNSGNDSIRKKINVGAGS----- 338
Db 1402 VDV-----ATLPGAANVQASVSNVAGNSQATHASVDATAPSVTINT 1445

Qy 339 -ADTAVNVAQLKEAVRLANRQ-----ITPKGDDSNRRVEKGLKTLITITGG----- 384
Db 1446 IATDDILNAAEAGSALTISGTSTAEAGQTVTVTLNGVYSGNVQADGWSVSVPTGDLAS 1505

Qy 385 -----AQTSAITDHNIGVQNGDGLKYLQALETLSLKKVYTNEL----- 423
Db 1506 LTASSTYVNASVSKARNSASATHN-----LTVDLAAPVVTINTVAGDDIINATEH 1556

Qy 424 -----TANE-KVTVGKRLITDKIGFTNDMNGIDSKPYLDKDTGIGHAGQ 468
Db 1557 GQAQIISGSATGATTGNVTSTVITGTYTT-----VLDANG-----NWSIGVPAS 1601

Qy 469 KIKTLTAG-----VVDODAATYQLKKVQNTAES-----ALQFTTVKKVQKNGNDANDSK 518
Db 1602 VISALAQGDVITITATVTDSDAGNSGTASHTVTVVALGAPVLAINITAVDDIINAAEKADLA 1661

Qy 519 IITVGNKPKDGTQVNTLKLKGENGVDVTTETNG--TVTFGLNQNGI-----TVGNSTL 571
Db 1662 I--TGTSNQAPAGTQI--VTLNGQN--YTTTADAGNSVTVPAKRSVALGEATYTVTAAT 1717

Qy 572 NNDGLSVKNTSNKQIQVAGDGIFF-----TD-----ISNSKPGAGIENTTRI 614
Db 1718 DADGNS-GSASHNVQVNTALPGVTINNVATDDIINAAEAGVEQITISQVGTGAAGDVTIV 1776

Qy 615 TRDIGFANNTGSLDANKPLTPTGINA-----GKELTNVQSAINPATNG----- 661

Db 1777 T---LGCATYTTATQANLSWSDVPASALQELNGELTISASVTNSVNGTNGTREITID 1833
Qy 662 -----QLDFMRLSTANTEKSGSAAT---KOLYNLSQVPLTFAGDT----- 700
Db 1834 ANLPLGRVDTVAGDDVVNIIIEHQALVITGSSSGLAAGSNVTLTINGQTYVAALADGTW 1893
Qy 701 ---GPNVTKK---LGEILKVKGGKTTAD-----DLTKNNI---GVVADSTNSL 740
Db 1894 SVGPADVNSAWPAGSVTIAASGTSAGNPVSVTHPTVVDLSAVASVINATADDVINA 1953
Qy 741 TVKLAKTSLDLDAVNTKTLTASDKVTVDGSGNNT-----AKLQNGD 780
Db 1954 EKGAAALTLSG---STSGVEAGQTVTVTFGKTYTSATVAANGSWSTSVPAAADMAALRGD 2009
Qy 781 LT--FSKQNTGATPATNSKTIQVD-----GLKFTDNN----- 810
Db 2010 ASAQASVSNVNGNSATTHAYSVDSAPTVTINTIAGDDILNAAEAGAALTTIGSSSTABA 2069

Qy 811 ---GIALDGTYYITKDKVGFAPKQDGLDKSKPYLDKDKL-----KVG----- 849
Db 2070 GQTVTVTLNGY-----TGTVQTDGWSVSVPSADLSTLTASNYTVNAAVSDRAGNPAS 2124
Qy 850 -----EVEITNGINAGGKAITGLSNTLT-----DATNATGHTVQLGIVDST--- 892
Db 2125 VHNHLLTVDTSPVVTINTVAGDDVINATEHAQAQIISGSATGAATGSTVTVTIGTNTFTT 2184

Qy 893 --DKTRAASIG-----DVLNAGFLKNGDKAKDFVS-----TYD 924
Db 2185 VLDASGNWSVGPASVSVSALANGTVTINASVTDAG---GNSGSATHQVTVNTGLPTTFN 2241

Qy 925 TV---DFINGNATTAKVTVYDGKSKVAYDVNVVDGTTIHLTG-----ADGNKNQIGVK 973
Db 2242 AISGDNILNADERQOPLTISGSGTGLATGQV---TVTLNGHNSYATTDASGN-----W 2292

Qy 974 TTTLTCTDKADGKAINFVSNGDDKALINAKDIADNL-----NTLAGE--IRNT 1020
Db 2293 TLTVPVSDLAALGOANYTVSASATSAAGNTASSQANLLVDSGLPDVTTINTVAGDDIINAA 2352

Qy 1021 KGTADTALQTFQVKKVKENGDDDDADDTITVQKDAKTQVNT-----LKLKGG 1068
Db 2353 EAGADQITISGVVTRAAA-----GDTVTVTLGGNTYATVQSNLSWSVSVPTADLQALGN 2406

Qy 1069 NGLDIQ---TNKQGTVTFGINTQSLKAGANNNTLNNN--GLSINKNTAGNEQIQ----- 1116
Db 2407 GDLTITASVTNANG-----NTGSETR---DITIDANLPLGRVDTVAGDDIVNSIEHQ 2457

Qy 1117 -----VQADGVKFAKVNNGVVGAGI--DGTTRI-----TRDEIGFA 1150
Db 2458 LVITGGSSGLNAGAVLTVTINSVAYSATVQADGWSVGIPAAANVSAMPAGPLTVEVDQGS 2517

Qy 1151 GTNGSLDKSKPH-----LSKQGINAGGKKITNIQSGETAQ-NSNDAVT--- 1192
Db 2518 SANNPVSVSHPTVVDLTAVAIISINTVASDDVINAEEKGTNLTLGSGTSGHESGQTVTVTF 2577

Qy 1193 GKKTY-----DLKT-----ELENKISSTAKTAQNSLHEFSVADEQGN 1229
Db 2578 GKKTYTASVAANGSWSVNVPADLTLPEGAANVQASVSSASGNSASATHAYSV-DASAP 2636

Qy 1230 NFTVSNPYSSVD---TSKTSQDVITFAG-----ENGITTKVN-KGVVRVIGIDQTKG---LTT 1278
Db 2637 TLTI-NTIASDDIILNAAEAGSPLTISGTSTAETQGTVTVTLNGATYTGTVQADGWSVSV 2695

Qy 1279 PKLTVGNNGKGIIVD---SONGONTITGLSNTLANVTNDKGSVRTTEQKKIIDEKDKR 1335
Db 2696 PTSNALGALNASYTVSATVNDKAGNPGSASHNLAVDTIAPVLIINTVAGDDIINDAEHAQ 2755

Qy 1336 AASIVDVLISAGFLQNGGEAVDFVS-----TYDTVNFADGNATTAKTVDYDDTSKTSKV 1389
Db 2756 AL-VISGTSAG-----GEAGDVVSVVLNGKTYTITLTDASGNWS----- 2792

Qy 1390 YDNVDDTTIEVKDKKLGVKTTTLTSTGTGA-----NKFALSNQATGDALYKASDIV 1441

Db 2793 -----VGVPAADVTLGSGAQTTITASVSDRAGNSDDASRTVTVLSAPV 2836
Qy 1442 AHLNLTSGD---IQTRAGASQANSSAGYVDADGNKVLYDSTDNKYQAKNDGTVDKTEV 1498
Db 2837 ISINTIAGDDVINATEKSGDLALS--GTSQDPAGTATITVTLNGQNYSATTDASGNMSVTV 2894
Qy 1499 AKDKLVAQAQTPDGTGLQAMNVKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDNKTKN 1558
Db 2895 P-----ASAVSALGEATYSVTASVTN-----AQGNSSTASHNVQVN 2930
Qy 1559 AAVTVGDLNVAQPTLTFAGDGTGTTAKLGETLTIKGGQDNTNKLDDNIGVVGAGDTGFT 1618
Db 2931 TALPGITINPVATDDIINASEAGSAQTISQVGTAAAGSTVTVEL-----GCKTVT 2981
Qy 1619 VKLAKDLTNVSNVAGGTKIDDKGVSFVDSGQAKANTPVLNGLDLGCKVSNVKGKT 1678
Db 2982 ATVQADLSNNVSPVPAWDQ-----ALNGELTIVNASVTNA-----VGNTGSQT 3024
Qy 1679 KD--TDAANVQQLNEVRNLLGL--GNAGNDNADGNQVNIAD--IKKDPNSGGSSNRTVIK 1732
Db 3025 RDIITIDA-----SLPGLRVDTVAGDD-----VNNLIEHAQAQVITGSSG--FA 3066
Qy 1733 AGTVLGGKGNNDT---EKLATGQVQVQVDKGNANGDLNVMVVKTKQKGSKKALLATYNA 1789
Db 3067 AGTALTIVINNQTYAATVLANGWSVGPAT-----DVSN-W---PAGTLNITVSGANS 3116
Qy 1790 AGQTNVLTN-----NPAAIDRIN--EOGIRF-----FHVNDGNEQFPVQ 1828
Db 3117 AGTQTSITRPLTVLDTAVAIMSNISITSDAINAAEKGAALTLSGSTSGVEAGQTVTVTFG 3176
Qy 1829 RNGIDS--SASGRHSV--AIGFQAKADGEAAVAIGROTQAGNQSTAIADNAQATGDQSI 1883
Db 3177 GKYTTTVAANGSWSTVPAADLAALRDGDASQV--RVTVN-----GNSATATHEYSV 3229
Qy 1884 -----AIGTNVNAKHSGA---ICDPSTVKADNSYSV--GNNNQFTDATQTDV 1927
Db 3230 DSAAPTVTINTIASDNIINASEAAAGVTVSGTSTAQTGTLTVTLNGTVNQT--VQTD- 3286
Qy 1928 FGVGN-NIIVTESNVALGSNSAISAGT--HAGTOAKSDG-TAGTT-----TTAG- 1974
Db 3287 ---GWSLTPASDLTALANNGYTLTATVSDLAGNIGSAGKGVTVDTTAPVIFSNFVAGD 3343
Qy 1975 -----ATGTVKG-----FAGQ-----TAVGAVSVGASGAERIQNV 2005
Db 3344 DVINNVEHIOAQIISATGAVAGDRILVITIAQQVVTSTDSAGMSVGPASV--ISGL 3401
Qy 2006 AAGEV--SATSTDVNGS-----QLYKATQGIANATNELDRIHQENKKNAGISSAM 2056
Db 3402 ADGTVTISATITDSAGNSSTQTHNVQVNTAAVSLSVSTISGDNLI---NAAEAG--SAL 3455
Qy 2057 AMASMPQAVIPGSMVTGGIAT--HNGQAVAVGLSKLSDNGQWFKI-----2102
Db 3456 TLSG-----TGTFATGTVTVTLNKGYSAT---IQSNGSWSNVPAADVAAALSDGT 3505
Qy 2103 -----NGSADTQGH 2111
Db 3506 STVVSASAQDSAGNSSTQTH 3527

RESULT 9
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
R;Accession: B85547
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; PMID:21074935; PMID:11206551
A;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA

A;Residue: 1-5188 <STO>
A;Cross-references: GB:AE005174; NID:G12513368; PIDN:AA054838.1; GSPDB:GNO0145; UWGP:Z06
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0615

Query Match 5.9%; Score 634; DB 2; Length 5188;
Best Local Similarity 22.0%; Pred.No.1.5e-16;
Matches 576; Conservative 290; Mismatches 928; Indels 830; Gaps 130;

Qy 35 TQGVGSVRTLSFARIAALAVLVIGATLN---GSAYAGIGISEADGG-KGGANARGDKSI 89
Db 1136 SGPFTLTINTVSGDDIINAAEIVVAQTIISQVGTGTAAGTAVTIVTIGNQVNAIVQSDLSW 1195
Qy 90 AI---GDIAQALGSGSIAIGDNKIVHNSNNNANIGAKSGNESIAIGGDVLASGHASIAI 146
Db 1196 SVSVPANVLQALNGELTISAS--LTNSANN-----TGTATHDVIDANFLGRLVDTV 1246
Qy 147 GSDDLVLKKEETVQOISELLPIIRGQKALNDIYQLADTNLQKYRRTHAQGHASTAV--- 201
Db 1247 AGDDVINSIEHTQAL-----VITGSS-----GLAAGAALTIVVINSVT 1284
Qy 202 -GAMSYAKGHFSNAFGRATAE---GTYSIAGVLGTATAKAASIAVGSNAQAIFFAATAV 257
Db 1285 YGATVLADGWSVGVFPVADVTNWPAGTVNIAVSGTNTAGTTTSI---SHPVTVDLAVAL 1341
Qy 258 -----GGSTQVNLNRGIALGFGSQ---VLQKDN---DVNA 286
Db 1342 TINTLTDDVINAAEKSGDLQSLGTTSGVEAGQTTIVFGKSYTTTVAADNTWGLTIPA 1401
Qy 287 ANVRAPVADPNQPIDNRYKATFKNGATDV-FSIGNSNGNDSIRRKIINVGAGS----- 338
Db 1402 VDV-----ATLPDGAANVQASVSNVAGNSQATTHAYSDATAPSTINT 1445
Qy 339 -ADTDVAVNVAQLKEAVRLANRQ-----ITFGDDSNNRVERKGLKTLITGG----- 384
Db 1446 IATDDIILNAAEAGSALTISGTSTAEAGQTVTVTLNGVNYSGNVQADGWSVSVPTGDLAS 1505
Qy 385 -----AOTSALTDHNGVQNGDGLKVQLAETLTSUKMYTTEL-----423
Db 1506 LTASSYTVNASVSDKARNASATHN-----LTVDLAAPVVTINTVAGDDIINATEH 1556
Qy 424 -----TANE-KVTVGKTRLTDTKIGFTNDMNGIDESKPYLDKDTGIHAGG 468
Db 1557 GQAQIISGSATGATGNTVSVTIGTITYT-----VLDANG-----NWSIGVPAS 1601
Qy 469 KITKLTAG-----VVDDDAATYQQLKKVNTAES-----ALQFTVVKYVDKNGNDANDSK 518
Db 1602 VISALAQQDVTITATVTDGAGNSGTASHTVTVALGAPVLAINITIAVDDIINAAEKGADLA 1661
Qy 519 IITVGNKPKDGTQVNTLKGKENGVDVVTETNG--TVTFGLNQNNGL-----TVGNSTL 571
Db 1662 I--TGTSNQAGTQI--TVTLNGQN-YTTTADAGSNMSVTPASRVSALGEATVTVTAAAT 1717
Qy 572 NNDGLSVKNTNSKQIOVGADGITE---TD-----INSKPGAGIENTTRI 614
Db 1718 DADGNS-GSASHNVQVNTALPGVTINVAVDIINAAEAGVEQTIISQVTTGAAGDVTIV 1776
Qy 615 TRDGIGFANNTGSLDANKPRLTPTTGINA---GGKELTNVQSAINPATNGG-----661
Db 1777 T---LGGATYVATVQANLSWSDVPASALQELNGELTISASVTVNSVGTNGTRETITD 1833
Qy 662 -----QLDFMRLSTANTEKSGSAATI---KDLNLSQVPLTFAGDT-----700
Db 1834 ANLPGLRVDTVAGDDVNNIIIEHQALVITGSSSGLAAGSNVTLITINGQTVYAAVLADGTW 1893
Qy 701 ---GNVTKK---LGEILKVKGKKTAD-----DLTKNNI---GVVADSDTNSL 740
Db 1894 SVGVPAVDVSNWAPGASVTTIAASGNSAGNPVSVTHPVTVVDLSAVAVSINAITADDVINAA 1953
Qy 741 TVKLAKTSLDLDAVNTKTLTASDKVTVDSGNNT-----AKLQNGD 780
Db 1954 EKGAALTLSG-----STSGVEAGQTVTVTGGKTSYATVAANGSWSTSVPAADWAALRDGD 2009

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: B89921

A;Status: preliminary

A;Molecule type: DN

A;Residues: 1-6713 <KUR>

A;RESIDUES: 1-0713 (KOR)
A;Cross-references: GB:BA000018; PID:q13701232; PIDN:BAB42527.1; GSPDB:GN00149

A: Experimental source: strain N315

C:Genetics:

A; Gene: ebhA

Query Match 5.8%; Score 622; DB 2; Length 6713;
Best Local Similarity 20.5%; Pred. No. 6e-16;
Matches 559; Conservative 317; Mismatches 970; Indels 876; Gaps 124;

QV 6 KVI FNKATGT FMAVAEYAKSHSTGGGSCATGOVGSVRTL SFARIAALAVLVIGATLNGSA 65

[illegible]

Db 39 ETILNKQTGP--NTAKTAVEQALNNVNSAKHALNGTQNLNNAKQAAITAINGASDLNQKQ 96

[illegible]

QY 66 YAGIGISEADGGKGGANARG-----DKSIAIGDIAQALGSQLAIGDNKTVHNSNNAN 119

[illegible]

Db 97 KDALK-AQANGAQRVSNANDVQRNATELNTAMGQLQHAIADKTNTLASSKYVNADSTKQN 155

Qy 120 -IGAKASGNESIAIGGDVLASGHASIAIGSDDLKKTETVQIQISELLPIRGQKALNDIY 178

[illegible]

Db 156 AYTTKVTNAEHIISGTPVTVTTPSEVTAAN-----QVNSAKQELNGDERLR--- 202

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

QY 179 QLADTNLQKYRRTHAQGHASTAVGAMSYAKGHFSNAFGTRATAEGTYS LAVGLTATAKAA 238

[illegible]

DB 203 -----VAKQNANTAI DALTQ-----LNT PQAK 225

04 239 SSI2VCSNA0A1GEF8ATAVGCSSTOYNI NPGI A1 GEGSSOVLKONDINIA ANB2AY-----A 283

QY 239 SSI AVGSNAQAIGFAAIAVGGSITQVNLNRGIALGFGSQVLQKUNDVNAANVRAY-----A 293

[illegible]

DD ZZO LREQQGQANKLEDVQS VQINGQSTNNAMKGL-----RDSIANEIIIVNASQNIIDA Z/Z

OV 294 PDDNOPIDNRYKATFKNGATDVFSIGNSNGNDSIRBKIIINVGAGSADTDVAVNO--LKE 351

[illegible]

Db 276 SPNNQ-----STYNSAVSNAKGIINQTNPT-----MDTSAITQATTQVNNAKNGLNG 323

CSC INTERNATIONAL TRADING COMPANY
1 YEAR CONTRACT AGREEMENT
MAY 10 9 47
22

Qy 352 AVR LANRQITFK-----GDDSNRNRVEKGLKLTITGGAQT 387

1D Lattice Chain

Diagram illustrating a 1D Lattice Chain with sites numbered 1 to 10. The chain is represented by a horizontal line with vertical segments at each site. The sites are labeled 1 through 10 from left to right.

Db 324 AENLRNAQNTAKQNLTLSHLTNNQKSAISSQIDRAGHVSEVTAAKNAATELNAQMGNLE 383

[illegible]

QY 388 SALT D H N I G V V Q N G ----- D G L K ----- V Q L A E T L ----- T S L K M V T T E N T L T A N 426

— — — — —

Db 384 QAIHDQN--TVKQGVNFTDADKAKRDAYTNAVSR AETILNKQTGANTSKQDVE----AAI 437

```
QY 1620 KL-----AKDLTNLSVNA-----GQTKIDD----- 1640
Db 1829 KQALNGQENLTAQNAKHLNGLSLTNAQDKAAKROGIEATHVETQONNADALN 1888
QY 1641 -----KGVSVFVDSGGQAKANTPVLSSANGLDLGGKVIS-----NVGKGT 1678
Db 1889 TANTNLKNGIQDQNTIKQGVNFTDAD-EAKRNA---YTNAVTAQAEQILNAQCPNTAKDG 1944
QY 1679 KDTDAANYQOL-----NEVRNLLGLGNAG----- 1702
Db 1945 VETALQNVQRAKNELNGQNVANAKTAKNALNLTSSINNAQKALKSQIEGATTVAGVN 2004
QY 1703 -----NDNAD-----GNQVNIADIKKDPNSGS 1724
Db 2005 QVSTWASELNTAMSNLQGINDEAATKAAQKYTEADRDQTAYNDAVTAATKILLDKTAGS 2064
QY 1725 SNRTVIK-----AGTVLGGKGN-----NTEKLATGG 1752
Db 2065 NDNKVAVEQALQRVNTAKTALNGDARLNEAKNTAKOOLATMSHLTNAQKANLQEIERGT 2124
QY 1753 VQGV-----DKQ-----NANGDLNVMVVKQKDGSKKA 1782
Db 2125 TVAGVQGIQANAGTLNQANMLQRQSTASKDATKSSDYQDANADLQNAV-----ND----- 2175
QY 1783 LLATYNAAGQNTYLTNNPAAEADIRNEQIRPFHVNDGQEPVVOGRNGIDSSASGKHSV 1842
Db 2176 --AVTNAEGIIS-ATNPEWNPDTINQASQV-----NSAKSALNGDEKL 2217
QY 1843 AIGFQ-AKADGSAVAIGRTOQAGNOSIAIGDNAQATGQDSIAIGTG-----NVVAG 1893
Db 2218 AAQKQAKSD-----IGRLTDLNNAQ-RTAANAEDVQAPNLAATAAKNKAATSLNTAMG 2270
QY 1894 KHSGAIGBPSTVKADNSYSVGNNOFTDATQDTPVGVGNITVTESSVALGSN---SAL 1950
Db 2271 NLKHALAEKDNTRKSVNY-----TDADQPKQAYDTAQAEAITNANGSANVETQV 2322
QY 1951 SAGTHAGTQAK-----KSDGTAGTPTTAGATGTVKGFA 1983
Db 2323 QALNLQNAKNDLNGDNKVAQAKETAKALASYNLNAQSTAASTQSDNATTVADVTA 2382
QY 1984 GQ-----TAVGAVSVGASGABERRIQNVAAAGEVSATSTDAVNGSQLYKATQGIANATNE 2036
Db 2383 AQNTANELNTAMGOLONGIND-----QNTVKQOVNFTDADQ-GKKDAY--TNAVTTNAQGI 2434
QY 2037 LDHRIHQENKANAGITSSAWAMASMPQAVIPGRSMVTG-----GIATH--NGQ--- 2082
Db 2435 LDKANGQNTKQ--VEAALNQVTTAKNALNGDANVROAKSDAKANLGTTLHLNNAQKOD 2492
QY 2083 -----GAVAV-GLSKLSDNGQ 2097
Db 2493 LTSQIEGATTVNGSVSKTKAQ 2514

RESULT 12
AB0480
probable invasin YPO3944 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0480
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0480
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3013 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC93406.1; PID:gl15981852; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3944
```

```
Query Match 5.6%; Score 598; DB 2; Length 3013;
Best Local Similarity 21.2%; Pred. No. 1.8e-15;
Matches 547; Conservative 314; Mismatches 984; Indels 736; Gaps 118;

QY 13 TCTFWAVAYAKSHSTGGSCATQGVSVRTLSFAR-----IAALAVLVIGATLNGSAYA 67
Db 519 THTISAVAYDAQGNASN-----RAVTSIEVTRPETWISHLATTIDNATANGIATN 569
QY 68 GIGISEADGGKGANARGDKSIAIGDIAQALGQSQIAIGDNKIVHNSNNNANIGAKASGN 127
Db 570 TVQATVTDGD-----GQPIIQGLINFVAFTQATLSTTEATG--ANGT 610
QY 128 ESIAIGDVLASGHASIAIGS-----DOLYLKKEVQOQISEL-LPIIRGQKALNDIYOLA 181
Db 611 ASYTLTHTVSGVSRVSVILGSSRSRVDVTFVADESTAEITAANLTVTTNDSVANG-----S 666
QY 182 DNLQKYRTHAQGHA-----STAVGAMSVAKGHFSNAGFTRATAGTYSVLAVGLTA 233
Db 667 DTNVVRKVTDAYTNAVANSQVIFSASNGATVIDQTVITNAEG-----IADST-----LTN 717
QY 234 TAKAASSIAGVSNQAIGFAATAVGGSTQ-----VNL-NRGIALGFGS---OVLQKNDVN 285
Db 718 TTAGSVWTTATLGGQSQQVDTTFKPGSTAAISLVKLADRAVADGIDQNEIQVLRDGTGN 777
QY 286 AA-----NVRA-----VAPDDNQPIDNRYKATFXN--CATDVFSI----- 318
Db 778 AVPNVPMISOADNGAIVVASTENTGVDGTINATFTNLRAGESVSVSPALVGMTMTMTF 837
QY 319 -----GNSGNDNS--IRRKIINVAGSADTDVAVNVAQLKEAVRLANR 358
Db 838 SADPRFVAVVSTLAAIDNNAKADGTDNVVRVAVWDANGNSVPGSVSTFDAGNGAVLAQNP 897
QY 359 QITFKGDDSN-----RVEKGLGKLT-----ITGGAQTSALTTHNIGVQN 400
Db 898 VYTDNGRYAENTLTNLALGTTTTCATTTVPQGVNTHFVAGAVDTITLTPVNGAVAN 957
QY 401 G---DGLKVLQAE-----LTSIKMV--TTENLTANEKVTVGKT-----RLTDDKIGFT 444
Db 958 GYNTSVQAVVSDSGNPNVTGATVVFSSNTAQVTTVIGTGVGDGIATLINTVAGTS 1017
QY 445 NDMNGIDESKPYLDKDTGTHAGGQKITLTA---GVVDDAATYQOLKKVNOTAESALQT 501
Db 1018 NVVATIDTVNANI--DTAFVAGAVATITLTPAVNGAVADGADT---NQVDALVE----- 1066
QY 502 FTVKVKDKNGN-----DANDSKLITVGKKNKPDG----- 530
Db 1067 -----DANGNPITGAUVVFSSANGATILSTMTNGVNGVASTLLTHTVAGTSNVVATVD 1120
QY 531 ---TVNTLKLKENG-VDVTTTETNGTTFGLNQNNGLTVGNSTLNN--DGLSVKNTSN 584
Db 1121 TVNANIDTTFVAGAVATITLTPVNGAVADGANSVQAVVSDSDGNPVTGAUVFSSAN 1180
QY 585 KOIQV-----GADGITFDISNKPAGIENNTTRITRDGIGFANNTGSLDANKPRLTP 637
Db 1181 ATAQITTVIGTTGADGIATATLTNTV--AGTSNVV-ATIDTVNANIDTAFVAGAVATITL 1237
QY 638 T-----GINAGGKELTN-----VQSALNPATNGQQLDFMNR-----LSTANTEKSSAAT 682
Db 1238 TAPVNGAVADGAD-TNQVDALVQDANGNAITGAUVVFSSANGADIATPTMTNGVNGVAST 1296
QY 683 IKDLYNLSQVPLTFAGDTGPNVTKKLGEILKVKGGKTTADDLTKNNIGVAVDSTDNLSLV 742
Db 1297 L-----LHTVAG-----TSNVVATIDTISANIDTA 1322
QY 743 KLAKTLSLDVAVNTKTLTASDKVTVSDGNNNTAKLQ-----NGD-----LTFESKONTGA 790
Db 1323 FVA-----GAVATITLTPAVNGAVADGADTINQVDALVEDANGNPITGAUVFSSAN-CA 1375
QY 791 TPATNSKITGVNGLKFT-----DNNGIALDGTYYITKDKVGFAGQDGLSKPYLD 842
Db 1376 TILSSTMTNGVNGVASTFLTHTVAGTSNVVATIGSVTENIDTAFVAGAVATITLTPAV-- 1433
QY 843 KDKLKVGEVE--ITTINGINA-----GGKAITGLSNTLTDATNATTGHTVQLGIVDSTDXT 895
```



```
Db 1434 -----NGAVADGVNTNSVQAVVSDSGNVTGATWFESSA-NATAQITTVIG-TTGADGI 1486
QY 896 RAASIGDVLNAGELNKNKGDAKPVSTYDVTDFINGNATTAKVTYDGGKASKVADVNVVDG 955
Db 1487 ATATLTNTVAGTSNV-----VATIDTV---NANIDTFFA--GELENIV--VSIIN 1530
QY 956 TTIIHLTGADGNKNGIKGVKTTTLTKTDAGDKAINFNSVNSGDDKA-LINAKDIADNLNTLA 1014
Db 1531 NNALANGADINIVEAFV-----TDRFGNGVANQSLIFGNGASIVGSSVTVTNLD--- 1580
QY 1015 GEIR-----NTKGPADTAL-----OTFQVKVK-----ENGDDND 1045
Db 1581 GRVRASATHVAGSSNTVIAISGAHQYARVTFVADVSTAQLKLTSLDNLQLANGKAGNI 1640
QY 1046 ADITVTKDAKTQV--NTLKLKGNGLDITKNKGIVTFGI-----NTQSGL----- 1091
Db 1641 AQALVT--DAHDNILLANQSVSFALDNGAVIESQGDASSASGIVLMRPNNTLAGMTTWTAT 1698
QY 1092 --KAGNNTLTNNGL-----SIKNTAGNEQIQVGADGVKFAKVNNGVVG-- 1133
Db 1699 LDSTGQETLETHFVAGKAASIEMTMTKDNVANNIDTNEVQVLVTDVDGNALGAVNVL 1758
QY 1134 -----AGIDG--TTRITRDEIGFAGTNGSLKSKPHLSKDGINA---GGKK 1174
Db 1759 TNSGMNITPNSVTTGSDGTATATLTHTLAGSLPINARIIDQ---VSKT-INATFIADAS 1813
QY 1175 ITNIQSBEIAQNSNDVATGGKIYD-----LKTELENKISSTAKTAQNSLHEF 1221
Db 1814 TAQLIAGDMFIINDQVANGQAVNAQVARDVSDVGNFPIKQTVFVLSLNGTITQYELDVT 1873
QY 1222 SVADEQGNFTVSNPYS-----SYDTSKTSDDVIT-----FAGENG 1256
Db 1874 SV--EGGVMTFTTLAGITNTVATVSSGSSRNIDTTFIADVTTHAIAASDLMIIVDDA 1931
QY 1257 ITTKVNGVVRVGLDQTKLTPKLTGVNNGKGIVIDSONGQNTITGLSNTLANVTNDK 1316
Db 1932 VADNLDKNEHARVTDKAGNLSQTVIFTSNGAAITTVNGISDGDGLTKATLTHTLAG 1991
QY 1317 GSVRTTEQGIKIIDEDKTRAA--SIVDVLGAFNLQNGEADVPVSFYDT---VNPADGN 1371
Db 1992 TSVVTRVGRNVQSKOTTFIADRTTATIRASDLTITRNALADGAVTNAARVITVDANGN 2051
QY 1372 ATTAKVTYDDTSKTSKVVDVNVDDTTFIEVKDKLGVKTTLTSTGTGANK-----FALS 1426
Db 2052 PVPSMEFVGYTSDNCALLTPSGMTDSS-----GTFSTFTHTTAGISKVTAIIVTMG 2103
QY 1427 NQATGDALVKASDIVAHLNTLSGDIQTAKASQNSS-----AGYVDADGN-----KVI 1475
Db 2104 ISQTKDAVFIAADRSTAHVS-----ELIVVKNDLSANNSDRNIVQAHIKDAHGNVVTGMNVN 2159
QY 1476 YDSTDNKVYQAKNDGTVDKTEKVAKDKLVAQAQTPDGTCLQAMNVKSVINKEQVNDANKQ 1535
Db 2160 FSATENTVLTANTVTT--NSQGVAEVTLRHNAPVTSATATV-----ATDLV 2204
QY 1536 GINEDNAFVKGL-----EKAASD-----NRTKNAAVTVGDLNVAQAQTPLTTFAGDTG 1581
Db 2205 GLTEDRVFVAGAGARIELFRINDGAVADGQTNRVEARVVDVSD-NLVPNSNVVFSADNG 2263
QY 1582 -----TTAKKLGLET-LTIKGGQDTNKLTDNNIGVVAGT-----DGTIVKLAKD--- 1624
Db 2264 GQLVQNDVQTDALGSAVTVSNINTGVTKVTVTADGVSASTTTTFFIADRDATLTLVDRFL 2323
QY 1625 LTN-----LNSVNAAGTKIDDKGSFVDSQOAKANTPVLISANGLDLGGKV 1670
Db 2324 ITHDNVANGVVENRVLHLVDANDNSVGVNFSAATNGASINASAITDINGFAIG--V 2381
QY 1671 ISNVGKGTQTDAAANVOQLNEVRNLGLGNAGNDN-----ADGNQVNIADIKDPNS 1722
Db 2382 LTNTLSGFSDVTVTLVTP-----CGTESLTVTPQFIADINTANIA----- 2421
QY 1723 GSSSNRIVIKATVILGGKGNDDTEKLATGVQGVGDKNANGDLSNVWK----- 1773
```

```
Db 2422 --TGDFVIIDDGAVANSVDANEVRVVT-----DNQGNAIAGYSVVFSSQNGATITTS 2472
QY 1774 --TQKQSGKALLATYNAAGTNYLTN-----NPAE 1802
Db 2473 GITVDGWSAKU-THIKAGESGILARLSRPMATVHTLMPYFIADVSTATLQLFNFP-- 2529
QY 1803 AIDRINEQGIIRFF---HVNDDGNQEPVVGQGRNGIDSSAGKHSVAIGFQAKADGAAVAIG 1859
Db 2530 -IPIADGVNQVFFVLGRVFDANQNFV--GGQQVAFSATNEVTLTESNGSISTPBGSVLLS 2586
QY 1860 -RQTAQGNQSIAGIDNAQAQTDQDSIAIGTGNVAGKHSIGAIG-----DPSTVKADNYSV 1913
Db 2587 VTSTQAGVHPi-----TCTLSNNYTDTFGAFAIANKNTAQSLTLMVV 2629
QY 1914 GNNQFTDATQTDVFGVGNNTITVESNSVALGNSAISAGTHAGTQAKK---SD----- 1964
Db 2630 DNNALADGVTRNQVRA---HVDSTGNSVA---DMAVTFTANRGAQUSKVTVLTDNNGDA 2693
QY 1965 -----GTAGTTTATAGATGTVKGFAGQTAAGAV--SVGASGAERRIQ-- 2003
Db 2684 VNTLTNSLVGVTVTAKLGTAGTPLTVDTVFTAGPLATLTLVTTVNNAFADNSATNTVQA 2743
QY 2004 -----NVAAGEVSATSDAVNGSOLYKATOGIANATNELDHRHONENKANAGISAM 2056
Db 2744 TLKDVSGNPITVGEVFAFA--ASNGATITATDGGVSNANG-----IVLATLNGTAGVSTVT 2797
QY 2057 A-----MASMPQAVIPGRSM-VTGGIATHNGQGAJA-----VGLSKLSDNGQWVFKINGS 2106
Db 2798 ATITLTETTTDTTFIAMKNLDVTVNGTTFNGDAGFTTGFVGAT-----FKVNSGG 2848
QY 2107 D 2107
Db 2849 D 2849

RESULT 13
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohca, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu-
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447

Query Match 5.5%; Score 590.5; DB 2; Length 2271;
Best Local Similarity 17.0%; Pred. No. 2.4e-15;
Matches 385; Conservative 435; Mismatches 973; Indels 469; Gaps 78;

QY 24 KSHSTGGSCATQGVSVRTLSPARIALAALVIVIGATLN-----GSAYAGIG 70
Db 267 KQYMTTSGNATYQSTGIVTLTQDAYSQKGAITLGRIDSNSKSFHFSKVNLGKRYEGHG 326
QY 71 ISEADGCGK-----GANARGDKSIAIGDIAQALGSQSIAGDNKIVHNSNNANI 120
Db 327 -----NGSDGIGFAPSPCVLGETGLNGAAVGIGLSNAFGKL-----DTHYNT-SKPS 375
QY 121 GAKAGNESIAIGDVLASGHASIAIGSDDLYLKKTVOQISSELLPIIROQKALNDIYQL 180
Db 376 AAKANADPSNVGGAGFAPVTTDSYGVATVTTSSSTADNAKL-----NVQPTNNTFQD 430
QY 181 ADTNLOKRYRTHAQGHASTAVGAMSVAKGHFSNAFGTRATAGTYSLAVGLTATAKAASS 240
```


R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A;Reference number: AD3252; PMID:11756688
A;Accession: AB3528
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2554 <R>
A;Cross-references: GB:AB008918; PIDD:AAU53389.1; PID:gl7984282; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMB10148
A;Map position: II
C;Keywords: hydrolase; serine proteinase

Query Match 5.3%; Score 568.5; DB 2; Length 2554;
Best Local Similarity 21.4%; Pred. No. 2.1e-14;
Matches 492; Conservative 271; Mismatches 925; Indels 607; Gaps 104;

Qy 202 GMSYAKGHPNAGTRATAGTYSVLAVGLTATAKAASSIAVGSNAQAIGFAATA----- 256
Db 20 GAIGFTK--AGAGSVLTGQNTYGVVTLGGILVTVMANAGVSGIG-QSTADPANL 75

Qy 257 -----VGGSTQVNLNRGIALGFGSVLQKDNVNAANVRAYAPDNPIDNRYK-- 305
Db 76 MLESGTRYTGS--VITDRGFTLVNGGPVAVIEVTGSGNL-AFSGLVTSFDDAGEFK 132

Qy 306 ----ATFKNGATDVFSIGNSGNDSIRRKIIINVG-----AGSADTDVAVNAQLKEAVRL-- 355
Db 133 GAGTLTFEUNGNDHIGATTVSSGGLTAVSLADGGQVSSGLKSGSDATNLILAGGALNYLG 192

Qy 356 ----ANRQITPKGDDNNRKEGLKGLTITGGQTSALTDHNGTVGVNGDGLKVLQALBT 411
Db 193 STTSDDRSFTLGGAGNSIGVANA-GTTLMSG--TAVGTD--GLTKLGGD----- 237

Qy 412 LTSKAMVTENLTANEKVTYKTRLTDDKI-----GFTNDMNGIDESKPYLKDGTGHHAGG 467
Db 238 --TLLSCTNYTGTAVNAGVLRAGSAQAFPSGLMTVNG-----ASLELGG 284

Qy 468 QKIT--KLTAAGVDDDDAATYGQKKNVQTAESALQPTVKKVQKNGNDANDSKIIIVGK 524
Db 285 YDITVSGILGAGTVDLGGNTL-----TSSGSAANSFTGKITGTGGFRTGGSTQTLSG 337

Qy 525 NKKPDGQVNTLKLKGENGVDVTTETNGTVTFGLNQ--NGLTVGNSTLNDGLSVKNTN 582
Db 338 CNS-DYTGKTTIAGNGLSVLCLNKGQAGSSNAPDLNVLNGLSTYGTNTV-TTD 395

Qy 583 SNKQIQVGADGITFDISNKPAGAGIENTTTRITRDGIGFANNTGSL---DANKPRLTPTG 639
Db 396 RGFTIQGGTGALSVTDAATLTFSG-----QVVGFGALQKRDGTGLVLMNSNYSYR-GGTS 449

Qy 640 INAGKELTNVQSAINPATNGQGLDFMRLSTANTEKSGSAATIKDL--YNLSQVPLTFA 697
Db 450 VDAG-----TLRAGSGAFGGSGMSLSN-----AAGAILDLGFDTSVTSLSGG 493

Qy 698 GDTGPNVTKLGEILKVGKGTATDLDL-----KNNIGV-----VADSTDNLSLT 742
Db 494 GALGNVALG-GATLTISSGNSNGTSYTGALTGTGNFVKNNGTQRLTGCCASSYGGSTTI 552

Qy 743 K-----LAKTSLDLDVNTKLTASDKVT--VDSGNNTAKLNGDLTFKQNTGATPATNSK 797
Db 553 NGGVLEVSCLADGGSVSSIGMSADADNLVINGGVLRVTGSGDSTDRQFTLGASGGNSIE 612

Qy 798 TIGVDGLKFTDNGI---ALDGTYYITKQVGPQKQDGLSKPKPYLDKXKLVGEVEIT 854
Db 613 SEGTAIILFTSNAAVTFAAANTAQTTLTLTAGTNTDDNERG-----AQLT 657

Qy 855 TNGINAGKALITGLSNT-----LTDATNATTHVTOL-GIVDSTDK-----TRAASIGDV 903
Db 658 NN-----GSGTSLTKTDTGTWFLTNSDSTYTG-VTKINGGVLSVDKLANGVLASSIGAS 711

Qy 904 LNAGFNKNGDAKDFVSTYDVTDFINGNATTAKVTYDYGKASKVAYDVNVVDGTTIHLTGA 963

Db 712 SSAASNL-----IIGNDSTLRYLGTGDTTDRILFTLASGLTYESSGS 753

Qy 964 DG-----NKNQIGVKTTLTKTDKAGDKAINFVSNGDDKALINAKDADNLTLAGERN 1019

Db 754 GAIVFTDTGOVALADNNQARTIALGGK-----NTGD-----NTLAGSIGD 793

Qy 1020 TKGTDATLQTFQVKVKENGDDDDADDTITVGDKAKTQNVNTLK--LKGKNGLDIQNK 1077

Db 794 A-GTGKTTLA-----KNDDGTGVLVTGNTTGTGPTNINKGLLKING-----GT 835

Qy 1078 DGTVTFGI-NTQSGLKAGNNNTLNNGLSLIKNTAGNEIQVCGAGVGFPAKVNNGVVGAG- 1135

Db 836 TGSLSLTDIVVTDGGLIFNRSDTLAYGGL-----ISSAGFVTQSGSGTTLTGCANSYTGATS 891

Qy 1136 -IDGTTTRITRDEIGFAG-----TNGSLDKSKPHLSKDGINAGGKKITNIQSEIAONSDA 1190

Db 892 VSAGTLLVNGDQSAATQTSVANGSI-----LGGSGIIGNVVVTD--GALAPGSGA 942

Qy 1191 -----VTGGKIYDLKTELENKIS-----STAKTAQNSL-- 1218

Db 943 GTLTINGSLALSAGSILSMQLGQAGVAGGALNDL-IEVKGNLTLDGLTLDVAETAGSGYCP 1001

Qy 1219 -----HEFSVADEQGNFTVSNPYSSYDTSKTSDDVITFAG-----ENGITTKVNGK 1264

Db 1002 GIYRLINVTGSLTDNGLDGLMPLNGAGAIQPAVAGQVNLLAGTGNFNFMDGDVGPKEFSA 1061

Qy 1265 VVRVGIOTKGLTTPKLTIVGNN-----GKGIVIDSQNGQNT 1301

Db 1062 V-----DGGNGTWQNSGNNMTDATGNINASYSDGAFATGTGTAGTVIDRSLGQVK 1114

Qy 1302 ITGLSNTLAN--VTNDKGSVTRTEOGKIIKDEDKTRA-----ASIVDVLSAGFNLQNGE 1354

Db 1115 AEGQPAIDSVAVTGDK--LELTGPQSTIRVGDGTGAGAAVIATINSULTQNTQLE-KTD 1171

Qy 1355 AVDFVSTYDVTNFADG--NATTAKVTYDDT--SKTSKVYDVNVDDTIEVKKDKLGVK 1409

Db 1172 AGTLVLT-GANSYTGTAINGGTIRISSDNLGVASSDISPDGALNTTANITADRAIIL 1230

Qy 1410 T-----TTLTSTGTGANKFALSNOATGDALVKASDIVAHLNLTLSGDIQTAKGASQ 1459

Db 1231 TGAGTLTLDASTLTLSSGPIISGTCALTKSGGTLL--SGTAH-----TGTTTTAGTFLQ 1284

Qy 1460 ANSSAGVVDADGNKV-----IYDSTDNKYQAKNDGTVDKTKE-----VAKDKLVAQA 1507

Db 1285 IGRGTSSIDGNIVNGALVDFDRAGTLAYTGSISGVTGLTKNGSSLTITMTGTSTYTGCT 1344

Qy 1508 QTPDGTLAQM--NVKSVIN-----KEQVNDANKKQGINEDNAFVKGLEKAASDNKTKN 1558

Db 1345 TVSAGTLALQAGGQIKGTASLTVDGGAEVLIDGSGSQFATGAGASVVG-----T 1393

Qy 1559 AAVTVGDLNVAQTPLTPLFAGDTGTAKKLGTLIKGGQDTNKLTDNNIGV--VAGT-- 1614

Db 1394 GTTVRDGTTASFDLSLTSTNATGNS-----TITVAG--SGSQMTQTGATFAGLAGTAT 1445

Qy 1615 ----DGFTV-----KLAKDLNLNSVNAAGTK-----IDDKGVS 1644

Db 1446 VDILDDGTMISSGASVFGGQLPMDATGQVTISAGSGQWTIANALYARRGSITVDDGGVV 1505

Qy 1645 FVDSGQAKANTPVLSANGLDLGGKVISNVGKGTQD--AANYQQLNEVRLNLGLGNAG 1702

Db 1506 TAGSAVIGYADTGINNPE-TDL---VVTGAGSRPETTGELAITNSAANAARSGSITADGG 1561

Qy 1703 NDNADGNQVNIADIKDPNSGSSNRTVIRKAGTVLGGKNNNDTEKLATGGVQGVVDKDN 1762

Db 1562 VVKVGGGALAMPGNVAVLINIGAAAGGSPAHAGTLDAG-----TVMVAGSQINFNHD-- 1614

Qy 1763 ANGDLSNVVKTKQDGSKKALLATYNAAGQTNLTNNPAAIDRINEGIRIFFHVNDGNQ 1822

Db 1615 ----DASTTFSATISGAS-----VSGSGGATLLTGNNSYAGLTITVTAGSLYI---DGDQ 1663

Qy 1823 EPVQGRNGIDSSASGKHSV-----AIGFOAKADGEAAVAIGRQTAQGNQSTA-----IG 1872

Db 1664 -----SMATGLTIVNPGTGGTGTIGDVTVASGGAINPGSPGMAPGTLLING 1712
QY 1873 DNAQATGD-QSIAIGTGNVAG-----KHSAGIDPSPVTKADNS 1910
Db 1713 DUTLASGSTQSFSGQANIPGGPLNDLINVGBDLVLAGTLQVDTSAAGTMDPGIYRFPNY 1772
QY 1911 YSVGNNGQFT-DATQTDVFE---GVGNNTIVTEGNSVAL-----GNSAISAG--- 1953
Db 1773 TGTLSQNAWTVNLPSPDFYQTSVAQQLNVNTAGLALRFWDGADPQNKNGKLEGGNGI 1932
QY 1954 -----THAGT-QAKSDGTAGTTTATAGATGVKGFAGQATVAGVAVS----- 1992
Db 1833 WQAFGSAPDNGNDWMTETGNTINAPFQDATEAVFTGEKGTVTVDSDSKAGINVGSIQFVTDG 1892
QY 1993 -----VGSAGAEIRLQNVAG---EVSATSDAVNGSOLYKATQG----- 2029
Db 1893 YIVNGDAINLVGASGSTIRVGGTGTGTVDASIDABITGASOLIKADMGTLLITGDSNY 1952
QY 2030 -----IANATNEL-----DRIHQENKANAG----- 2051
Db 1953 TGGTKITGGTLQVAKDSALGTRTCELLIDGGTLNTTADMTIDRSGITDQAGTLDIDTGT 2012
QY 2052 --ISSAMAMSPQAYI---PGRSMVTGGIATHNGQGAVALGSLSD--NGQWVFKING 2104
Db 2013 LKIDGVLSCAG---AFVKITGAGRLELAGDDHTYNGDASIASGTALTALGALGGTMNVGIDG 2069
QY 2105 SADTQGHVGAACVAG 2119
Db 2070 RLEATGRVGAATNSG 2084

RESULT 15
A97859
190-Kda cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: A97859
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 233, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; PMID:21442074; PMID:11557893
A:Accession: A97859
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03811.1; PID:g15620410; GSPDB:GN00173
C:Genetics:
A:Gene: rompA

Query Match 5.1%; Score 550.5; DB 2; Length 2021;
Best Local Similarity 21.5%; Pred. No. 7.6e-14;
Matches 511; Conservative 243; Mismatches 811; Indels 811; Gaps 116;

QY 14 GTFMAVAYAKSHSTGGSCATQGVSVRTLSFARIALAVLVIGA-----T 60
Db 88 GDYITADVADHIITAINVADTTPIG---LNTAQTVVGSIVTGNLLPVITTAGKSLT 143
QY 61 LNSAVAGTIGISEADGGKGANARGDKSIAIGDIA-----QALGSQSIAGDNKIVHNSN 115
Db 144 LNS-----NNADAANHFQAPADNYTGLGNIALGGANAALIIQSAAPAKITLAGNIN 195
QY 116 NNANIGAKASGNESIAIGG-DVLASGHASIAIGSDLLYLLKKEVQOISELLPIIRQKAL 174
Db 196 GGGIITVTDAAINGIIGNTNALATVNVGAGIATLEGAIIKATTKITNAASVL----- 249
QY 175 NDIYQLADTNLQYRRETHAQQGHASTAVGAMSAYAKHFSNAFGFR-----ATAEQTYSILAVG 230
Db 250 -----TLTNVNAVLTGALDNTTGVNDVNVGLNGLALGALSOVTG 285
QY 231 LTATAKAASIAVGSNAQAGFAATVAGGSTQVNLNRGIALFGSOVLQK---DNDVNAA 287
Db 286 NIGNTNALATISVAGKATILGGA--VIKATTTKLTADNASAVTFTNPVWVGTGALDNTGNA- 342

QY 288 NVRAYAPDDNQPIDNRKYATFKNGATDVPFISGNSGNDSSIRKLIINVAGSAD----- 340
Db 343 -----NNGIVTFTGDSVTGNGNTNA-----LATISVAGKATILGGAIIK 383
QY 341 -----TDAVNAQLKEAVRLA-----NRQITFKGDDNN-----RVEKG 374
Db 384 ATTTKLTADNASAVTFTNPVVTGADTNGNANGIVTFTGDSVTGNGINTNALATISVG 443
QY 375 LGKTLITIGG---AQTSALTDNHIGVVQNGDGLKVQLAETLTSJKMVT-----TENJTAN 426
Db 444 AGRA-TLGGAIIKATTTKLTADNASAV-----TFTNPVWVGTGALDNTGNAN-N 488
QY 427 EKVTVGKTRLTDKIGFTNDMGIDSKPYLDKDTGHAG-----GQIKTLTAGVWDDA 482
Db 489 GI VFTFGDSVTGNGINTNALAT-----SVGAKATILGGAIIKATTTKLTADNA 537
QY 483 ATYGQLKKVNOT-----AESALQTF--VKKVDKNGNDANDSKIITVCKNKP-DGTQ 532
Db 538 SAVTFTNPVVTGADTNGNANGIVTFTGDSVTGNGINTNALATISVAGKATILGGA 597
QY 533 VNTLKLKGVGVDTTETNGTVTFGL-----NONNGLT--VGNSTLNNDGLSVKNTNSNK 585
Db 598 IKATTTKLTADNASAVTFTNPVVTGADTNGNANGIVTFTGNSVTG---NIGNTNALA 654
QY 586 QIOVGA-----DG-----ITFTDISNSKPGAGIENTTITRDIIGIFPANN 624
Db 655 TVNVGAGIATLEGAVIKATTTKLTNAASVLTNVNAVLTGA-IDNTGTG--DNVGVNL 711
QY 625 TGSLED-----ANKPRLTPTGINAGGKELTNVQSAI-----NPAATGGQL 663
Db 712 NGALSQVTGNGINTNALATISVAGKATILGGAIVKATTTKLTADNASAVTFTNPVVTGAI 771
QY 664 DFMRLSTANTEKSGSAIYKDYLNLSQVPLPAGDTGPNVTKLGE-----I 711
Db 772 D-----NTGNA-NGGIA-----TFTGDS--TVTGNIGNTALATVNVGAGL 809
QY 712 LKVGKGTITADDLTKNIGVWADSTNSLTVLAKTLSDLDVNTKLTASDKVTVDSG- 770
Db 810 LRVOG-----GVKSNNTIN-----LTD-----NASAVTFTNPVVTGAI 843
QY 771 NNTAKLQNGDLTFSKQ--TGATPATNS-KTIGVDGLKFTDNNNGIALDGTFTYTKQVGF 827
Db 844 DNTGNANGIVTFTGDSVTGNGINTNALATISVAGKATILGGAIIKATTTKLTADNA--- 900
QY 828 AKQDGLSKSPYLDKDKLVGVEVITTINGINAGGKAITGLSNTLTDATNATTHVTVLQ 887
Db 901 ----SAVFTNPV-----VVTGALDNTGNANGIV-----TFTGDSVTG 936
QY 888 IVDSTDKTRAASIGDVLNAGENLKNG--DAK--DFVSTYDVTDFINGNATTAKTAVYDGG 943
Db 937 NIGNTNALATVNVG-----AGVTLQAGGSLDANNIDF-GARSTLEF-NG-----PLDGG 983
QY 944 ASKVAYDVNVDDGTTIHLTG--ADGNKNQGVKTTTLTKTDAKDKAINFVNSGD----- 996
Db 984 GNAIPY-----YFKGAIANGNAILNVNTKLTAYHLTIGTVAEINIAGNLPAID 1034
QY 997 ----DKALNAKDIADNLNTLAGETIRNTKGTADTALQTFQVKKVKGNGDDDN---DADTI 1049
Db 1035 ASAGDVITILNAQDI--HPRAL-----DSALVLSNLTGSGVN-----NILLAADLV 1077
QY 1050 TVGKDAKTQVNTLKLKGNGLDIOTNKDGTVTFGINTQSGLKAGN--NTTLNNGLSIK 1107
Db 1078 AGVDEGTVVD-----GGVNGLINISNVAGAA-----RNIGDVGGKFTLTIYNAVIT 1128
QY 1108 NTAGNEIQVAGDGVKFAKVNNGVWVAGIDGTTTITRDEIGFAGTNGSDKSKPHLSKDG 1167
Db 1129 D-----DYNLEGIQNVLIN--NADFTSSTAFNAGTIQINDATYTTIDANNGLN--- 1175
QY 1168 INAGGKKITNIQSGETAQS--ND-AVTGGKIYDLKTELENKISSPATAQNSHEFSVA 1224
Db 1176 IPAGNIKFAHADAQLILQNSGNDRTITLGANIDPDNDDEGIVILNSVTA----- 1225
QY 1225 DEQGNFTVSNPYSSYDTSKTSDDVITFAGENGITTKVNGVVRVGDIDQTKLITPKLTVG 1284

Db 1226 ---GKLTIAAGKTFGGAHKLQDIV-FKGE-----GDFGTAGTTF----- 1261
Qy 1285 NNNKGIVIDSQONQITG---LSNTLANVTNDKGSVRITTEQOKLIKDEKTRAASIVD 1341
Db 1262 --NTNTNIVLD-----ITGQLEGATTANVLFKDAVOLQTQTG----- 1296
Qy 1342 VLSAGFNLOQNGEAVDFVSTYDVTNFDAGNATTAKVTYDDTSKTSKVYDVNVDQTTIEV 1401
Db 1297 -----NIGG-----FLDPNAKNGTIVTLN-----WNVVAGT----- 1322
Qy 1402 KDKLGVKTTTLTSTGT-----GANKFALSNQATGDALVKASDIVAHLNLTLSGDIQTAKGA 1457
Db 1323 -----VKNTGGTNGTLLVLGASNL--NRVNGIAMLKVG-----AGNVTIAKG- 1363
Qy 1458 SQANSSAGYVDADGNKYIYDSTDNKYIOAKNDGTVDTKEVAKDKLVAQAQTPDGTLAQM 1517
Db 1364 --GNVKIGEIGTGTNTL--TLPAHF--KLTSINKT-----GQQAALKL 1401
Qy 1518 NVKSVINKEQVNDANKKQGINEDNAFVKGLEKAAASDNKTNAAVTVGDLNAAVAQOTPLTFA 1577
Db 1402 N-----FMNG-----GSVSGVVGTAANSV 1420
Qy 1578 GDTGTT-AKKLGEFLTITG-----GQDTNKLTDNNIGVVGAGTDGFTVKLAKDLT----- 1626
Db 1421 GDITTAGATSPASSVNAKGTATLGTTTSAFTFTNTGAVTLAKGSITTSFAKNVTATSFVA 1480
Qy 1627 NLNSVNAG-----GTKIDDKGV-----SFVDS-----SGQAKANTPVL 1659
Db 1481 NSATINFGNLAFNSNITGSGTTLTLGANQVYTYGTSGFTDTLTLNTTFDGAAGSGGNIL 1540
Qy 1660 SANG--LDLGGKVTSNVKGKTKTDAAVVOQLNEVRNLLGLGNAGNDNADGNQVNIADIK 1717
Db 1541 IKSGSTLDLSG--VSNLALVVTAT-----NFDMN----- 1567
Qy 1718 KDPNSGSSNRTVTKAGTVLGGK-----NND-----TEKLIATG 1751
Db 1568 ---NISPDTKVTISAEATAGGLKPTPKENVKITTINNDNRFPVDFPDASTLTLFAEDIAAG 1624
Qy 1752 GVQGVGDKDGNANGDLSNV-----WVKTKQDGSKK-----ALLATYNAAGQT 1793
Db 1625 -----VIDEDFAPGGPLANI PNAANIKKSLELMEDAPNGSDARQAPNFGMLTPLOEADAT 1680
Qy 1794 NYLTNN---PAAIDIRINEQGI-----RPFHVNDGNOEPVVOGRNGIDSSAS 1837
Db 1681 THLMQDVVKPSTIAAVNVQVVASNISNITALNARMKVKQAGNKGPVSSGDEDM----- 1736
Qy 1838 GKHSVAIGFQAKADGEAAVAIGRQTAQGNQSIAGDNAQATG-----DQSIATGTG 1888
Db 1737 -----AKFGAWISPFVGNATQKMCNSIS-GYKSDTTGTTGTFDGFVSDDLVGLLA 1785
Qy 1889 NVVACK-----HSGAIGDPSTVKAD-----NSYSVGNNOFTDA----- 1922
Db 1786 YTRADTDIKLNKNTGDKRKNVESNIYSLYGLYVPYENLFVEAIIASYSNKKIRKSRVI 1845
Qy 1923 ---TQTDVFGVGNNTITVTSNSVAL-----GSNSAISAGTHAGTQAKKSDGTAGT 1969
Db 1846 ATTLETGVGYQTANGVKSESYTGQLMAGYTYMMSSENILTPLAGLRYSTIKDKSYKETGT 1905
Qy 1970 TTTAGATGTGKFAQTAVGAVSGVCSGABERRIQNVAAGEVSAITSDAINGSQLYKATQ- 2028
Db 1906 TY---QNLTVKGNKNTFDGLIGAKVS-----SNINYNEIVLT-----PELYAMVDY 1949
Qy 2029 GIANATNELDHRHQENKANKAGISSAMAMASMPQA 2064
Db 1950 AFKNKVSALDARLQ-----GMTAPLPTNSFKQS 1977

```

Query Match      66.0%; Score 138; DB 2; Length 1964;
Best Local Similarity 69.8%; Pred.No. 2.4e-07;
Matches 30; Conservative 3; Mismatches 10; Indels 0; Gaps 0
QY 1 IGISEADGGGGGNARGDKSIAGDIAQAALGSQSIATCDNKIV 43
      :::::
Db 69 IATGDSGLNGNARGADEKSIAGSIACATGSQSIATGDNKAV 111

```

```

RESULT 2
Q8KQM8
ID Q8KQM8 PRELIMINARY; PRT; 2314 AA.
AC Q8KQM8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hemagglutinin.
GN HAG.
OS Moraxella catarrhalis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Moraxella.
OX NCBI_TaxID=480;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=O12E;
RX MEDLINE=22112901; PubMed=12117964;
RA Pearson M.M., Lafontaine E.R., Wagner N.J., St Geme J.W. III,
RA Hansen E.J.;
RT "A hag Mutant of Moraxella catarrhalis Strain O35E Is Deficient in
RT Hemagglutination, Autoagglutination, and Immunoglobulin D-Binding
RT Activities."
RL Infect. Immun. 70:4523-4533(2002).
DR EMBL; AY07638; AAL78285.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 9.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
SQ SEQUENCE 2314 AA; 238336 MW; 3PFD721F67D4E2FD CRC64;

Query Match 43.1%; Score 90; DB 2; Length 2314;
Best Local Similarity 52.3%; Pred. No. 0.13;
Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;

QY 4 SEADGGKGGANAR---GDKSIAIGDIAQALGSGSIAIGDNKIV 43
DB 2040 AKADGEAAVAIGRQTQAGNSIAIGDIAQALGSGSIAIGDNV 2083

RESULT 3
Q8PGSO
ID Q8PGSO PRELIMINARY; PRT; 2190 AA.
AC Q8PGSO;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XAC3546
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo C., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,

```

```

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AB012003; AAM38389.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 53.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2190 AA; 204715 MW; 1BEA011869A3C2CA CRC64;

Query Match 37.1%; Score 77.5; DB 16; Length 2190;
Best Local Similarity 45.2%; Pred. No. 3.6;
Matches 19; Conservative 7; Mismatches 11; Indels 5; Gaps 1;

QY 2 GISEADGGK-----GGANARGDKSIAIGDIAQALGSGSIAIG 38
DB 1048 GESEAGAQAQSTALGAAAGAYGDSGLAVGALSQAQSGSESTAMG 1089

RESULT 4
Q9F3X6
ID Q9F3X6 PRELIMINARY; PRT; 1299 AA.
AC Q9F3X6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mapa protein (Hsf).
GN MAPA OR HSF 2 OR PM1570.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Pm70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AJ277635; CAC14202.1; -.
DR EMBL; AB006194; AAK03654.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 10.
DR Pfam; PF05662; HIM; 3.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1299 AA; 130963 MW; 8BCCE0EB66CDB428 CRC64;

Query Match 36.8%; Score 77; DB 16; Length 1299;
Best Local Similarity 45.7%; Pred. No. 2.3;
Matches 21; Conservative 7; Mismatches 12; Indels 6; Gaps 2;

QY 1 IGISEADGGKG-----GANARGDK--STAIGDIAQALGSGSIAIGDN 40
DB 109 IGFGATNDGNTVAIGAKSKSKAASTAIGDNKAUDNQAIAGQN 154

RESULT 5
Q8PCQ5
ID Q8PCQ5 PRELIMINARY; PRT; 2351 AA.

```

Q9PCQ5; 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Outer membrane protein.
GN XADA OR XCC0658.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; Pubmed=12024127;
RA da Silva A.C.R., Ferrero J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.I., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spingola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
RL EMBL; AE012164; AAM39974.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep Hag; 59.
DR Pfam; PF05662; HIM; 7.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2351 AA; 220261 MW; C5482A38C940DA11 CRC64;
Query Match 36.8%; Score 77; DB 16; Length 2351;
Best Local Similarity 60.7%; Pred. No. 4.4;
Matches 17; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 11 GGANRGDKSTAIGDIAQALGSQSIATG 38
DB 863 GQSNAAAGDESGALGWEAQAGDQGIALG 890
RESULT 6
Q9LAX0 Q9LAX0 PRELIMINARY; PRT; 1328 AA.
AC Q9LAX0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Xanthomonas campestris pv. pelargonii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=91612;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xpel-1;
RX Kim J.F., Zumoff C.H., Beer S.V.;
RT "An alanine-, glycine-, and serine-rich protein and a putative serine
RT protease of Xanthomonas campestris pv. pelargonii."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083618; AAF63394.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR

```

Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Rotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
[2]
RN
SEQUENCE FROM N.A.
RA STRAIN=O157:H7 / RIMD 0509952;
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
[3]
RN
SEQUENCE FROM N.A.
RA STRAIN=RIMD 0509952;
RC STRAIN=RIMD 0509952;
RA Makino K., Yutsudo C.H., Yokoyama K., Kubota Y., Kimura S.,
RA Shinagawa H.;
RT "O157 specific gene similar to H. influenzae adhesin gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR ENBL; AB005587; AAC58749.1; -
DR ENBL; AP002566; BAB37903.1; -
DR ENBL; AB036416; BAB87814.1; -
DR PIR; AB6036; AB6036.
DR PIR; H91188; H91188.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM-
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 13.
DR Pfam; PF05662; HIM; 12.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein; Complete proteome.
SQ
SEQUENCE 1588 AA; 160150 MW; B2BA4E06E28DEDC CRC64;
Query Match 33.78; Score 70.5; DB 16; Length 1588;
Best Local Similarity 32.88; Pred. No. 17; Indels 19; Gaps
Matches 19; Conservative 9; Mismatches 11;
Qy 4 SEADG-----GKG-----GANARGDKSIAIGDIAQALGSQSIAGDNKI 42
Db 281 NKADGVDAIALNGSQSRGLNTIALGTASNATGDKSLAGNSGANGINSVALGDSI 338
RESULT 9
Q98HJ2 PRELIMINARY; PRT; 1953 AA.
ID Q98HJ2 AC
Q98HJ2 AC
Q98HJ2 AC
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein ml12848.
GN ML12848.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OC NCBI_TaxID=381;
RN [1]
RP
SEQUENCE FROM N.A.
RA STRAIN=MAFF303099;
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).

```



```

DE Putative inner membrane protein.
GN STM3691.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sunderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008871; AAL22550.1; -.
DR InterPro; IPR008640; Hep Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 10.
DR Pfam; PF05662; HIM; 13.
DR Pfam; PF03895; Yada; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1461 AA; 147837 MW; 91C59A87E7282254 CRC64;

Query Match 32.5%; Score 68; DB 16; Length 1461;
Best Local Similarity 43.3%; Pred. No. 30;
Matches 13; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 13 ANARGDKSIAIGDIAQAALGSQSIAGDNKI 42
DB 186 ATASGDDSAAFNGAKAIGTNSVALGSGSV 215

RESULT 12
Q8FCB2 PRELIMINARY; PRT; 1778 AA.
AC Q8FCB2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative adhesin.
GN C4424.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016768; AAN82860.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 14.
DR Pfam; PF05662; HIM; 15.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 1778 AA; 177678 MW; 98564AA3A797DA20 CRC64;

Query Match 32.5%; Score 68; DB 16; Length 1778;
Best Local Similarity 53.8%; Pred. No. 37;

Matches 14; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 15 ARGDKSIAIGDIAQAALGSQSIAGDN 40
DB 170 ADGKTIAGNTAKAYEIMSIAGDN 195

RESULT 13
Q8F3X5 PRELIMINARY; PRT; 2712 AA.
AC Q8F3X5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MapB protein (Hsf).
GN MAPB OR HSF_1 OR PM0714.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RA Henderson I.R., Nataro J.P., Cappello R., Stein C.;
RT "Evolutionary origins of the autotransporter proteins.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AJ277636; CAC14203.1; -.
DR EMBL; AE006108; AAK02798.1; -.
DR InterPro; IPR008640; Hep_Hag.
DR InterPro; IPR008635; HIM.
DR InterPro; IPR005594; Yada.
DR Pfam; PF05658; Hep_Hag; 14.
DR Pfam; PF05662; HIM; 10.
DR Pfam; PF03895; Yada; 1.
KW Complete proteome.
SQ SEQUENCE 2712 AA; 276154 MW; 3F5579D6F32FFA3D CRC64;

Query Match 32.5%; Score 68; DB 16; Length 2712;
Best Local Similarity 57.7%; Pred. No. 59;
Matches 15; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 17 GDSKSIAGDIAQAALGSQSIAGDNKI 42
DB 324 GESSVAIGDKAVSRGEASIAICKNAI 349

RESULT 14
Q9FDA0 PRELIMINARY; PRT; 1265 AA.
AC Q9FDA0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative outer membrane protein Xada.
GN XADA.
OS Xanthomonas oryzae (pv. oryzae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=64187;
RN [1]_
RP SEQUENCE FROM N.A.
RA Ray S.K., Rajeshwari R., Sonti R.V.;
RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
RT that is involved in virulence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288222; AAG01335.1; -.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:32:39 ; Search time 0.799546 Seconds
(without alignments)
2800.358 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGKGANARGDKS.....GDIAQALGSQSTAIGNKIV 43

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.5	31.3	434	1 YADA_YERPS	P10858 yersinia ps
2	63.5	30.4	455	1 YADA_YEREN	P31489 yersinia en
3	62.5	29.9	725	1 MASZ_PSEFL	O05137 pseudomonas
4	61	29.2	377	1 RPOA_CHLMU	O091n4 chlamydia m
5	60	28.7	278	1 T2D7_DROME	Q22722 droeophila
6	59.5	28.5	104	1 CH10_SCHPO	O59804 schizosacch
7	59	28.2	401	1 YK03_CAEEL	P34291 caenorhabdi
8	58	27.8	710	1 EFG_BUCBP	P59451 buchnera ap
9	57.5	27.5	403	1 CK05_HUMAN	Q9uhr6 homo sapien
10	57.5	27.5	676	1 ICP0_HSVBK	P29836 bovine herp
11	57	27.3	2946	1 NBEA_HUMAN	Q8afp9 homo sapien
12	56.5	27.0	245	1 YH77_ARCFU	Q28497 archaeglob
13	56	26.8	377	1 RPOA_CHLTR	P46449 chlamydia t
14	55.5	26.6	498	1 NDDD_ALCXH	P94212 alcaligenes
15	55.5	26.6	698	1 EFG_VTBCH	Q9kuz7 vibrio chol
16	54.5	26.1	728	1 MASZ_BRUME	O8yir3 brucella me
17	54.5	26.1	728	1 MASZ_BRUSU	Q8fz50 brucella su
18	54	25.8	348	1 ID12_LACPL	Q89w06 lactobacill
19	54	25.8	1287	1 SK12_YEAST	P35207 saccharomyc
20	53.5	25.6	331	1 MACS_BOVIN	P12624 bos taurus
21	53.5	25.6	442	1 GLMD_RHIME	O87392 rhizobium m
22	53.5	25.6	699	1 EFG_HAEIN	P43925 haemophilus
23	53.5	25.6	699	1 EFG_VIBPA	Q87145 vibrio para
24	53.5	25.6	699	1 EFG_VIBVU	Q86c08 vibrio vuln
25	53.5	25.6	700	1 EFG_PASMU	P57938 pasteurella
26	53.5	25.6	703	1 EFG_ECOLI	P02996 escherichia
27	53.5	25.6	703	1 EFG_SALTY	P26229 salmonella
28	53.5	25.6	1043	1 DSG1_BOVIN	Q03763 bos taurus
29	53	25.4	219	1 PGNE_ECOLI	P77366 escherichia
30	53	25.4	507	1 FLIC_SALBE	Q06968 salmonella
31	53	25.4	507	1 FLIC_SALON	Q06974 salmonella
32	53	25.4	595	1 PRIM_CHLTR	O84799 chlamydia t
33	52.5	25.1	107	1 YBJQ_ECOLI	P75819 escherichia

34 52.5 25.1 329 1 HEM2_MYCLE P46723 mycobacteri
35 52.5 25.1 698 1 EFG1_SHEON Q8ek71 shewanella
36 52.5 25.1 699 1 EFG_COKBU Q83es7 coxiella bu
37 52.5 25.1 702 1 EFG_BUCAI P57593 buchnera ap
38 52.5 25.1 702 1 EFG_BUCAP Q8k948 buchnera ap
39 52.5 25.1 702 1 EFG_YERPE Q8zjb3 yersinia pe
40 52.5 25.1 703 1 EFG2_PSEPK Q88fi4 pseudomonas
41 52.5 25.1 705 1 EFG_WIGBR Q8dh2 wigleswort
42 52 24.9 152 1 RL9_CHLTE Q8kam4 chlorobium
43 52 24.9 232 1 GIDB_COREF Q8fsv1 corynebacte
44 52 24.9 270 1 NUP1_PENCI P24289 penicillium
45 52 24.9 270 1 NUP3_PENSQ P24504 penicillium

ALIGNMENTS

RESULT 1
YADA_YERPS
ID YADA_YERPS STANDARD; PRT; 434 AA.
AC P10858;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Invasin precursor (Outer membrane adhesin).
GN YADA OR YOPA OR INVA OR YOPI.
OS Yersinia pseudotuberculosis.
OG Plasmid pIB1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]_TaxID=633;
RP SEQUENCE FROM N.A.
RC STRAIN=YPIII;
RX MEDLINE=88302441; PubMed=3043229;
RA Rosqvist R., Skurnik M., Wolf-Watz H.;
RT "Increased virulence of Yersinia pseudotuberculosis by two
independent mutations."
RL Nature 334:522-525 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YPIII;
RX MEDLINE=89343638; PubMed=2761389;
RA Skurnik M., Wolf-Watz H.;
RT "Analysis of the yopA gene encoding the YopI virulence determinants
of Yersinia spp."
RL Mol. Microbiol. 3:517-529 (1989).
CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
SURFACE.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X13883; CAA32088.1; -
CC PIR; S04534; S04534.
CC InterPro; IPR008126; Adhesion.
CC InterPro; IPR008640; Hep_Hag.
CC InterPro; IPR008635; HIM.
CC InterPro; IPR005594; Yada.
CC Pfam; PF05658; Hep_Hag; 4.
CC Pfam; PF05662; HIM; 1.
CC PRINTS; PR01756; OMADHESIN.
CC Plasmid; Virulence; Signal; Outer membrane.
KW

FT SIGNAL 1 25
PT CHAIN 26 434 INVASIN
SQ SEQUENCE 434 AA; 45054 MW; E2C55FB12B183D4 CRC64;

Query Match 31.3%; Score 65.5; DB 1; Length 434;
Best Local Similarity 46.5%; Pred. No. 2.4;
Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGGANARGDK--SIAIGDIAQ-----ALGSQSTAIIGNKI 42
| ||||| ||||| | : ||||| | :
DB 94 GAGGLNARAKDPYSIAIGATAEAAKPAAVAVGSGSIATGVNSV 136

RESULT 2

ID YADA_YEREN STANDARD; PRT; 455 AA.
AC P31489;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Invasin precursor (Outer membrane adhesin).
GN YADA OR YOPA OR INVA OR YOPL.
OS Yersinia enterocolitica.
OG Plasmid pYV.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=6471/76 / Serotype O:3;
RX MEDLINE=95020586; PubMed=7934875;
RA Tamm A., Tarkkanen A., Korhonen T.K., Kuusela P., Toivanen P.,
RA Skurnik M.;
RT "Hydrophobic domains affect the collagen-binding specificity and
RT surface polymerization as well as the virulence potential of the yada
RT protein of Yersinia enterocolitica";
RL Mol. Microbiol. 10:995-1011(1993).

CC -1- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS. THIS
CC PROTEIN IS AN ADHESIN FORMING A FIBRILLAR MATRIX ON THE CELL
CC SURFACE.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; X13882; CAA32086.1; -;
CC PIR; S04912; S04912.

CC InterPro; IPR008126; Adhesion.

CC InterPro; IPR008640; Hep Hag.

CC InterPro; IPR008635; HIM_

CC InterPro; IPR005594; Yada.

CC Pfam; PF05658; Hep Hag; 4.

CC Pfam; PF05662; HIM; 1.

CC Pfam; PF03895; Yada; 1.

CC PRINTS; PR01756; OMADHESIN.

CC Plasmid; Virulence; Signal; Outer membrane.

FT SIGNAL 1 25 INVASIN.

FT CHAIN 26 455

SQ SEQUENCE 455 AA; 47136 MW; AC12EF68C657DAC0 CRC64;

Query Match 30.4%; Score 63.5; DB 1; Length 455;
Best Local Similarity 44.2%; Pred. No. 4.3;
Matches 19; Conservative 6; Mismatches 9; Indels 9; Gaps 2;

QY 9 GKGG--ANARGDKSIAIGDIAQ-----ALGSQSTAIIGNKI 42
| | | | | | | | | | | : | | | | | :
| | | | | | | | | | | : | | | | | :

Db 59 GAGGLNASAKGIHSIAIGATAEAAKGAAVAVGAGSIATGVNSV 101

RESULT 3

ID MASZ_PSEFL STANDARD; PRT; 725 AA.
AC O05137;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Malate synthase G (EC 2.3.3.9).
GN GLCB OR FC2.4.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=L6.5;
RA Artgenave F.M., Delecu M., Vilagines R., Danglot C.;
RT "A functional glyoxylate bypass is mandatory for utilization of
RT alkanes by Pseudomonas fluorescens";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC CoA.
CC -1- PATHWAY: Glyoxylate bypass; second step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; Y11998; CAA72726.1; -;
CC HSP; P37330; 1D8C.

CC InterPro; IPR001465; Malate synthase.

CC InterPro; IPR006253; Malate synthG.

CC Pfam; PF01274; Malate synthase; 1.

CC TIGRFAMS; TIGR01345; malate syn G; 1.

CC Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.

FT ACT_SITE 340 340 CATALYTIC BASE (BY SIMILARITY).

FT ACT_SITE 631 631 CATALYTIC ACID (BY SIMILARITY).

SQ SEQUENCE 725 AA; 78861 MW; 0C3326FB67C9B381 CRC64;

Query Match 29.9%; Score 62.5; DB 1; Length 725;
Best Local Similarity 40.7%; Pred. No. 8.8;
Matches 22; Conservative 2; Mismatches 17; Indels 13; Gaps 3;

QY 3 ISEADG---GKGGANARGDKSIAIG-----DIAQALG-----SQSTAIIGNKIV 43
| | | | | | | | | | | : | | | | | :
| | | | | | | | | | | : | | | | | :

Db 148 ISEADGAEGKGYKVRGDKVIAFAFAFLDEAAPLSAGSHVDSGTGYKIADGKLI 201

RESULT 4

ID RPOA_CHLMU STANDARD; PRT; 377 AA.
AC Q9PJN4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (RNAP alpha

DE subunit) (Transcriptase alpha chain) (RNA polymerase alpha subunit).

GN RPOA OR TC0794.

OS Chlamydia muridarum.

OC Bacteria; Chlamydiae; Chlamydiales; Chlamydia.

OX NCBI_TaxID=83560;

RN [1]

NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150630; PubMed=7545910;
RA Kokubo T., Gong D.W., Wootton J.C., Horikoshi M., Roeder R.G.,
RA Nakatani Y.;
RT "Molecular cloning of Drosophila TFIID subunits.";
RL Nature 367:484-487(1994).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94037099; PubMed=8221891;
RA Goodrich J.A., Hoey T., Thut C.J., Admon A., Tjian R.;
RT "Drosophila TFIID40 interacts with both a VP16 activation domain and
RT the basal transcription factor TFIIB";
RL Cell 75:519-530(1993).
[3]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: TAFs are components of the transcription factor IID
CC (TFIID) complex that are essential for mediating regulation of RNA
CC polymerase transcription (by similarity).
CC -!- SUBUNIT: TFIID is composed of TATA binding protein (TBP) and a
CC number of TBP-associated factors (TAFs).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the TAF2G family.

CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer
CC entities requires a license agreement (See <http://www.isb-sib.ch/announc>
CC or send an email to license@isb-sib.ch)

Best Local Similarity 42.4%; Pred. No. 12;
Matches 14; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 8 GKGGANARGDKSIAIGDIAALGQSIAIGDN 40
Db 339 GGGSGGAPGSQVPMTEAQQIAVQIQARDN 371

RESULT 8

EFG_BUCBP
ID EFG_BUCBP STANDARD; PRT; 710 AA.
AC P59451;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Elongation factor G (EF-G).
GN FUSA OR BBP470.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola";
RL Proc. Natl Acad. Sci. U.S.A. 100:581-586(2003).
CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
the nascent protein chain from the A-site to the P-site of the
ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC EF-G/EF-2 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR HAMAP; AE014017; AAO27176.1; --
DR HAMAP; MF 00054; -- 1.
DR InterPro; IPR0004540; EF-G.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000640; EFG C.
DR InterPro; IPR009022; EFG III V.
DR InterPro; IPR005517; EFG IV.
DR InterPro; IPR004161; EFTU D2.
DR InterPro; IPR005225; Small GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00679; EFG C; 1.
DR Pfam; PF03764; EFG IV; 1.
DR Pfam; PF00009; GTP EFTU; 1.
DR Pfam; PF03144; GTP EFTU D2; 1.
DR PRINTS; PR00315; ELONGATNFACT.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small GTP; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 710 AA; 79363 MW; 037BA37B190557FD CRC64;

Query Match 27.8%; Score 58; DB 1; Length 710;
Best Local Similarity 40.5%; Pred. No. 28;
Matches 15; Conservative 6; Mismatches 10; Indels 5; Gaps 2;

QY 13 ANARGD-KSIAIGDIAALGQSIAIGD-----NKIV 43
Db 374 ANKEELKEVRAGDIAAALGKSVTTGDTCLDPNNVV 410

RESULT 9

CK05_HUMAN
ID CK05_HUMAN STANDARD; PRT; 403 AA.
AC Q9UHR6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Protein C11orf5 (Protein PON).
GN C11ORF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20069078; PubMed=10602999;
RA Lemmens I.H., Farnebo F., Piehl F., Merregaert J., Van de Ven W.J.M.,
RA Larsson C., Kas K.;
RT "Molecular characterization of human and murine c11orf5, a new member
of the FAUNA gene cluster";
RL Mamm. Genome 11:78-80(2000).
CC -!- TISSUE SPECIFICITY: Low expression in most tissues; highly
expressed in testis.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF119497; AAF23591.1; --
DR Genew; HGNC:1177; C11orf5.
DR MIM; 604575; --
DR InterPro; IPR007529; Znf HIT.
DR Pfam; PF04438; zf-HIT; 1.
SQ SEQUENCE 403 AA; 42883 MW; 2384BF806CC40E71 CRC64;

Query Match 27.5%; Score 57.5; DB 1; Length 403;
Best Local Similarity 41.9%; Pred. No. 18;
Matches 13; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

QY 9 GKGGANARGDKSIAIGDIAALG---SQSIA 36
Db 236 GEGPTNQGYTLAALGDLAQTGLRKARKQAVA 326

RESULT 10

ICP0_HSVBK
ID ICP0_HSVBK STANDARD; PRT; 676 AA.
AC P29836;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE Trans-acting transcriptional protein ICP0 (P135 protein) (IER
2.9/ER2.6).
GN ICP0.
OS Bovine herpesvirus type 1 (strain K22).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31519;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219360; PubMed=1313901;
RA With U.V., Fraefel C., Vogt B., Vlcek C., Paces V., Schwyzler M.;
RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
are 3' coterminal and encode a putative zinc finger transactivator

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```

RT protein.";
RL J. Virol. 66:2763-2772 (1992).
CC -!- PTM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL
CC ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY
CC CASEIN KINASE II.
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84464; AAA46061.1; -.
DR DR PIR; A38209; EDBE22.
DR DR HSSP; P28990; LCHC.
DR DR InterPro; IPR001841; Znf ring.
DR DR Pfam; PF00037; Zf-C3HC4; 1.
DR DR SMART; SM00184; RING; 1.
DR DR PROSITE; PS00518; ZF_RING_1; 1.
DR DR PROSITE; PS00089; ZF_RING_2; 1.
DR KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW DNA-binding; Early protein; Repressor; Phosphorylation.
FT ZN_FING 13 52 RING-TYPE
FT DOMAIN 284 331 ASP/GLU-RICH (ACIDIC)
SQ SEQUENCE 676 AA; 67701 MW; 9BB0683C9BFC65D CRC64;
Query Match 27.5%; Score 57.5; DB 1; Length 676;
Best Local Similarity 44.1%; Pred. No. 31;
Matches 15; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

QY 2 GISEADGKGGANARGD-KSIAIGDIAQALGSQS 34
Db 120 GGSSEAGGGAGGAEEAAGAGAGAGAAAGRA 153

RESULT 11
NBEA HUMAN
ID NBEA HUMAN STANDARD; PRT; 2946 AA.
AC Q9NFP9; Q9HCW8; Q9NSU1; Q9NW98; Q9Y6T1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurobeachin protein (lysosomal trafficking regulator 2) (BCL8B
DE protein).
DE NBEA OR LYST2 OR BCL8B OR KIAA1544.
DE Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC TISSUE=Brain, and Spleen;
RX MEDLINE=22150869; PubMed=12160729;
RA Dyonin V.G., Chaganti S.R., Dymova K., Palanisamy N., Murty V.V.V.S.,
RA Dalla-Favera R., Chaganti R.S.K.;
RT "BCL8 is a novel, evolutionarily conserved human gene family encoding
RT proteins with presumptive protein kinase A anchoring function.";
RL Genomics 80:158-165 (2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=restis;
RA Dueterhoeft A., Lauber J., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 606-1118 FROM N.A.
RC TISSUE=Embryonic head;
RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,

```

```

RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1919-2946 FROM N.A. (ISOFORM 1).
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281 (2000).
RN [5]
RP SEQUENCE OF 2428-2946 FROM N.A. (ISOFORM 1).
RA Tchernev V.T., McMurtre E.B., Nguyen Q.A., Mishra V.S.,
RA Barbosa M.D.F.S., McIndoe R., Kingsmore S.F.;
RT "Identification of LYST2, a brain-specific member of the Chediak-
RT Higashi syndrome gene family.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RX X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 2150-2563.
MEDLINE=22220051; PubMed=12234919;
RA Jogl G., Shen Y., Gebauer D., Li J., Wiegmann K., Kashkar H.,
RA Kroenke M., Tong L.;
RT "Crystal structure of the BEACH domain reveals an unusual fold and
RT extensive association with a novel PH domain.";
RL EMBO J. 21:4785-4795 (2002).
CC -!- FUNCTION: Binds to type II regulatory subunits of protein kinase A
CC and anchors/targets them to the membrane. May anchor the kinase to
CC cytoskeletal and/or organelle-associated proteins (By
CC similarity).
CC -!- SUBUNIT: Interacts with RII subunit of PKA (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8NFP9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NFP9-2; Sequence=VSP_050538, VSP_050539;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Predominant in many brain structures. Also
CC expressed at medium levels in spleen, thymus, prostate, testis and
CC ovary. Low level expression is seen in heart, kidney, pancreas,
CC skeletal muscle and intestine.
CC -!- DOMAIN: RII-alpha binding site, predicted to form an amphipathic
CC helix, could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer (By
CC similarity).
CC -!- SIMILARITY: Belongs to the neurobeachin family.
CC -!- SIMILARITY: Contains 1 BEACH domain.
CC -!- SIMILARITY: Contains 5 WD repeats.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a stop
CC codon in position 762.
CC -!- CAUTION: Ref.5 sequence differs from that shown due to a
CC frameshift in position 2900.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF467288; AAM53531.1; -.
DR EMBL; AL137748; CAB70903.1; -.
DR EMBL; AK001059; BAAS1485.1; ALT_SEQ.
DR EMBL; AB046764; BAB13370.1; -.
DR EMBL; AF072371; AAD41633.1; ALT_FRAME.
DR PDB; 1MI1; 27-SEP-02.
DR Genew; HGNC:7648; NBEA.
DR MIM; 604889; -.

```



```

DR EMBL; AE004124; AAF93534.1; -.
DR PIR; C82332; C82332.
DR HSSP; P13551; 1ELO.
DR TIGR; VC0361; -.
DR HAMAP; MF_00054; -, 1.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR009022; EFG_III_V.
DR InterPro; IPR005517; EFG_IV.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF03764; EFG_IV; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 698 AA; 76927 MW; F44FDA4DB8FC4ECC CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 698;
Best Local Similarity 46.4%; Pred. No. 53;
Matches 13; Conservative 4; Mismatches 10; Indels 1; Gaps 1;

QY 13 ANARGD-KSIAIGDIAQAIGSQSIAIGD 39
Db |||:| |||||:|:|
366 ANKRDEIKRAGDIAAAGLKDVTGD 393

```

Search completed: September 21, 2004, 23:43:19
Job time : 2.79955 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 23:24:53 ; Search time 4.42676 Seconds
(without alignments)
2744.569 Million cell updates/sec

Title: US-09-813-214A-1

Perfect score: 209

Sequence: 1 IGISEADGGKGGANARGDKS.....GDIQAALGSQSTAIGNKIV 43

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	209	100.0	43	2	AAW32274
2	209	100.0	43	4	AAE00703
3	209	100.0	2122	6	ABU08784
4	209	100.0	2123	4	AAE00701
5	190	90.9	40	4	AAE00705
6	121	57.9	24	2	AAW32276
7	121	57.9	24	4	AAE00702
8	90	43.1	1992	2	AAW04505
9	90	43.1	1992	4	AAE69137
10	90	43.1	1992	4	AAE69133
11	90	43.1	2047	4	AAE69134
12	90	43.1	2053	4	AAE69135
13	90	43.1	2139	6	ABP71294
14	90	43.1	2314	4	AAE69136
15	89	42.6	1946	6	ABU35023
16	81	38.8	3073	6	ABU21223
17	76	36.4	2265	6	ABU17199
18	76	36.4	2504	6	ADA34534
19	70.5	33.7	1588	7	ADC01413
20	69.5	33.3	270	6	ABU19868
21	69	33.0	1129	6	ABU20124
22	68	32.5	1099	6	ABU20296
23	68	32.5	1400	6	ABU45394
24	68	32.5	1461	6	ABU47415
25	68	32.5	1778	4	ABB52677

ALIGNMENTS

RESULT 1

AAW32274

ID AAW32274 standard; peptide; 43 AA.

XX AAW32274;

XX AC

XX 08-MAY-1998 (first entry)

XX DE

XX M. catarrhalis outer membrane protein (OMP)-106 peptide fragment 1.

XX KW

XX Outer membrane protein-106; OMP106; vaccine; immune response;

XX KW

XX cytotoxic antibody; Moraxella catarrhalis.

XX OS

XX Moraxella catarrhalis.

XX PN

XX WO9741731-A1.

XX PD

XX 13-NOV-1997.

XX PF

XX 28-APR-1997; 97WO-US007679.

XX PR

XX 03-MAY-1996; 96US-00642712.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Tucker K, Plosila L;

XX WPI; 1997-558601/51.

XX Outer membrane protein, OMP106, of Moraxella catarrhalis - used in

XX vaccines for producing immune responses against M. catarrhalis.

XX Claim 9; Page 23; 78pp; English.

XX This is a peptide fragment of a novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating cultivar. The peptide fragment can specifically bind to an antibody that binds the OMP106 polypeptide. The antibody is a cytotoxic antibody which mediates complement killing of M. catarrhalis. The OMP106 polypeptide, and its peptide fragments can be used in vaccines and antigenic compositions. They can also be used for producing an immune response in an animal against M. catarrhalis

XX Sequence 43 AA;

XX Query Match

XX Best Local Similarity 100.0%; Score 209; DB 2; Length 43;

XX Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISEADGGKGGANARGDKSIAIGDIAQAALGSQSIAGDNKIV 43
 |||||
 Db 1 IGISEADGGKGGANARGDKSIAIGDIAQAALGSQSIAGDNKIV 43

RESULT 2
 AAE00703
 ID AAE00703 standard; peptide; 43 AA.
 XX
 AC AAE00703;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE N-terminal #1 of M. catarrhalis outer membrane protein-106 (OMP106).
 XX
 KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.
 XX
 PN US6214981-B1.
 XX
 PD 10-APR-2001.
 XX
 PF 12-NOV-1997; 97US-00968685.
 XX
 PR 03-MAY-1996; 96US-00642712.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L, Tillman UF;
 XX
 DR WPI; 2001-281002/29.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 PS Claim 2; Col 31; 49pp; English.
 XX
 CC The present sequence is N-terminal of haemagglutinating Moraxella
 CC catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a
 CC therapeutic and prophylactic vaccine against M. catarrhalis infections of
 CC mammals. It is used for diagnosis of bacterial infections and as reagents
 CC for clinical or medical diagnosis of M. catarrhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarrhalis. It is also used as a probe to identify the
 CC presence of M. catarrhalis in biological specimens and to identify other
 CC bacteria that encode a polypeptide related to M. catarrhalis OMP106.
 CC OMP106-derived polypeptides are used as ligands to detect antibodies
 CC elicited in response to M. catarrhalis infections and also as immunogens
 CC for inducing M. catarrhalis-specific antibodies which are useful in
 CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis
 XX
 SQ Sequence 43 AA;
 Query Match 100.0%; Score 209; DB 4; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1.9e-19;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISEADGGKGGANARGDKSIAIGDIAQAALGSQSIAGDNKIV 43
 |||||
 Db 1 IGISEADGGKGGANARGDKSIAIGDIAQAALGSQSIAGDNKIV 43

RESULT 3
 ABU08784
 ID ABU08784 standard; protein; 2122 AA.
 XX
 AC ABU08784;

XX 28-MAY-2003 (first entry)
 DT Moraxella catarrhalis outer membrane protein, OMP106.
 DE
 KW Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;
 KW vaccine.
 XX
 OS Moraxella catarrhalis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..69
 FT Protein /label= Signal_sequence
 FT Region /label= Mature_OMP106
 FT /label= N-terminal sequence of mature OMP106
 FT /note= "Specifically claimed in claim 9"
 FT Misc-difference 779..880
 FT /note= "Encoded by ATCTAAGGC"
 XX
 PN US2002177200-A1.
 XX
 PD 28-NOV-2002.
 XX
 PF 20-MAR-2001; 2001US-00813214.
 XX
 PR 03-MAY-1996; 96US-00642712.
 PR 12-NOV-1997; 97US-00968685.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L;
 XX
 DR WPI; 2003-328486/31.
 DR N-PSDB; ABX93525.
 XX
 PT Novel outer membrane polypeptide of Moraxella catarrhalis, useful for
 PT producing an immune response in an animal, and as ligands to detect
 PT antibodies elicited in response to Moraxella infections.
 XX
 PS Claim 33; Page 28-34; 51pp; English.
 CC The invention relates to an isolated or substantially pure outer membrane
 CC polypeptide of Moraxella catarrhalis (OMP106 polypeptide). The
 CC polypeptide and its peptide fragments are useful for producing an immune
 CC response in an animal, and as ligands to detect antibodies elicited in
 CC response to M. catarrhalis infections. The polypeptide and its peptide
 CC fragment are further useful as active ingredients in vaccines against M.
 CC catarrhalis infections. The polypeptide is useful in immunoassays to detect
 CC an antibody binding the polypeptide is useful in passive immunisations
 CC M. catarrhalis in biological specimens, and also in passive immunisations
 CC against M. catarrhalis infections. An antibody binding the polypeptide is
 CC also useful to facilitate isolation and purification of the polypeptide
 CC and its peptide fragment, and as probes for identifying clones in
 CC expression libraries that have inserts encoding the polypeptide and
 CC peptide fragment. An antibody binding the polypeptide is also useful to
 CC diagnose M. catarrhalis infections. The polynucleotide encoding the
 CC polypeptide is useful as a probe to identify the presence of M.
 CC catarrhalis in biological specimens by hybridisation or PCR amplification
 CC and also to detect other bacteria that might encode a polypeptide related
 CC to M. catarrhalis OMP106. The polypeptide, an antibody binding the
 CC polypeptide, its peptide fragment and the polynucleotide encoding the
 CC polypeptide are useful as reagents for clinical medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis, as well as
 CC host defence mechanisms. The present sequence represents the amino acid
 CC sequence of M. catarrhalis outer membrane protein (OMP)-106
 XX
 SQ Sequence 2122 AA;
 Query Match 100.0%; Score 209; DB 6; Length 2122;
 Best Local Similarity 100.0%; Pred. No. 1.6e-17;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISADGGKGGANARGDKSIAIGDIAQALGQSQAIGDNKIV 43
 DB 69 IGISADGGKGGANARGDKSIAIGDIAQALGQSQAIGDNKIV 111

RESULT 4
 AAE00701
 ID AAE00701 standard; protein; 2123 AA.

XX AAE00701;
 AC AAE00701;
 XX
 XX 02-JUL-2001 (first entry)
 XX
 DE Moraxella catarrhalis outer membrane protein-106 (OMP106).
 XX
 KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.
 XX
 PN US6214981-B1.
 XX
 PD 10-APR-2001.
 XX
 PF 12-NOV-1997; 97US-00968685.
 XX
 PR 03-MAY-1996; 96US-00642712.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L, Tillman UF;
 XX
 DR WPI; 2001-281002/29.
 XX
 DR N-PSDB; AAD04029.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 PS Claim 7; Col 53-64; 49pp; English.
 XX
 CC The present sequence is haemagglutinating Moraxella catarrhalis outer
 CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic and
 CC prophylactic vaccine against M. catarrhalis infections of mammals. It is
 CC used for diagnosis of bacterial infections and as reagents for clinical
 CC or medical diagnosis of M. catarrhalis infections and for scientific
 CC research on the properties of pathogenicity, virulence and infectivity of
 CC M. catarrhalis. It is also used as a probe to identify the presence of M.
 CC catarrhalis in biological specimens and to identify other bacteria that
 CC encode a polypeptide related to M. catarrhalis OMP106. OMP106-derived
 CC polypeptides are used as ligands to detect antibodies elicited in
 CC response to M. catarrhalis infections and also as immunogens for inducing
 CC M. catarrhalis-specific antibodies which are useful in immunoassays to
 CC detect M. catarrhalis in biological specimens. Cytotoxic antibodies are
 CC useful in passive immunisations against M. catarrhalis

Query Match 100.0%; Score 209; DB 4; Length 2123;
 Best Local Similarity 100.0%; Pred. No. 1.6e-17;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISADGGKGGANARGDKSIAIGDIAQALGQSQAIGDNKIV 43
 DB 69 IGISADGGKGGANARGDKSIAIGDIAQALGQSQAIGDNKIV 111

RESULT 5
 AAE00705
 ID AAE00705 standard; peptide; 40 AA.

XX AAE00705;
 AC
 XX 02-JUL-2001 (first entry)
 DT
 XX N-terminal #2 of M. catarrhalis outer membrane protein-106 (OMP106).
 DE
 XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 OS Moraxella catarrhalis.
 XX
 PN US6214981-B1.
 XX
 PD 10-APR-2001.
 XX
 PF 12-NOV-1997; 97US-00968685.
 XX
 PR 03-MAY-1996; 96US-00642712.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Tucker K, Plosila L, Tillman UF;
 XX
 DR WPI; 2001-281002/29.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 PS Example; Col 31; 49pp; English.
 XX
 CC The present sequence is N-terminal of haemagglutinating Moraxella
 CC catarrhalis outer membrane protein-106 (OMP106). The OMP106 is used as a
 CC therapeutic and prophylactic vaccine against M. catarrhalis infections of
 CC mammals. It is used for diagnosis of bacterial infections and as reagents
 CC for clinical or medical diagnosis of M. catarrhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarrhalis. It is also used as a probe to identify the
 CC presence of M. catarrhalis in biological specimens and to identify other
 CC bacteria that encode a polypeptide related to M. catarrhalis OMP106.
 CC OMP106-derived polypeptides are used as ligands to detect antibodies
 CC elicited in response to M. catarrhalis infections and also as immunogens
 CC for inducing M. catarrhalis-specific antibodies which are useful in
 CC immunoassays to detect M. catarrhalis in biological specimens. Cytotoxic
 CC antibodies are useful in passive immunisations against M. catarrhalis

Query Match 90.9%; Score 190; DB 4; Length 40;
 Best Local Similarity 100.0%; Pred. No. 4.7e-17;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IGISADGGKGGANARGDKSIAIGDIAQALGQSQAIGD 39
 DB 2 IGISADGGKGGANARGDKSIAIGDIAQALGQSQAIGD 40

RESULT 6
 AAW32276
 ID AAW32276 standard; peptide; 24 AA.
 XX
 AC AAW32276;
 XX
 XX 08-MAY-1998 (first entry)
 DT
 DE M. catarrhalis outer membrane protein (OMP)-106 partial peptide fragment.
 XX
 KW Outer membrane protein-106; OMP106; vaccine; immune response;
 KW cytotoxic antibody; Moraxella catarrhalis; primer; probe.
 XX
 OS Moraxella catarrhalis.

XX WO9741731-A1.
 PN
 XX
 PD 13-NOV-1997.
 XX
 XX 28-APR-1997; 97WO-US007679.
 XX
 PR 03-MAY-1996; 96US-00642712.
 XX
 XX (ANTE-) ANTEX BIOLOGICS INC.
 PA
 XX Tucker K, Plosila L;
 PI
 XX WPI; 1997-558601/51.
 XX N-PSDB; AAT86522.
 XX
 XX Outer membrane protein, OMP106, of Moraxella catarrhalis - used in
 PT vaccines for producing immune responses against M. catarrhalis.
 PT
 XX Disclosure; Page 58; 78pp; English.
 PS
 XX This is a partial sequence of an amino terminal peptide fragment of a
 CC novel outer membrane protein-106 (OMP106) polypeptide. The OMP106 is an
 CC outer membrane polypeptide of Moraxella catarrhalis, an haemagglutinating
 CC cultivar. The encoding DNA can be used as a 5' primer for PCR
 CC amplification of a full length OMP106 DNA. The DNA fragment can also be
 CC used as a probe for screening M. catarrhalis genomic libraries for OMP106
 CC polypeptide coding sequences. The encoded peptide fragment can
 CC specifically bind to an antibody that binds the OMP106 polypeptide. The
 CC antibody is a cytotoxic antibody which mediates complement killing of M.
 CC catarrhalis. The OMP106 polypeptide, and its peptide fragments can be
 CC used in vaccines and antigenic compositions. They can also be used for
 CC producing an immune response in an animal against M. catarrhalis
 CC
 XX Sequence 24 AA;
 SQ

Query Match 57.9%; Score 121; DB 2; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGGKGGANARGDKSIAIGDIAQ 28
 |||||
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24

RESULT 7
 AAE00702
 ID AAE00702 standard; peptide; 24 AA.
 XX
 AC AAE00702;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE Internal fragment of N-terminal outer membrane protein-106 (OMP106).
 XX
 XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 KW bacterial infection; immunogen; cytotoxic; antibiotic;
 KW passive immunisation.
 XX
 XX Moraxella catarrhalis.
 OS
 XX US6214981-B1.
 PN
 XX 10-APR-2001.
 PD
 XX 12-NOV-1997; 97US-00968685.
 PF
 XX 03-MAY-1996; 96US-00642712.
 PR
 XX (ANTE-) ANTEX BIOLOGICS INC.
 PA
 XX Tucker K, Plosila L, Tillman UF;
 PI
 XX

DR WPI; 2001-281002/29.
 DR N-PSDB; AAD04030.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer membrane
 PT protein-106 polypeptide, useful for diagnosis of bacterial infections and
 PT as vaccine against Moraxella catarrhalis infection of mammals.
 XX
 XX Example; Col 43-44; 49pp; English.
 XX
 XX The present sequence is the internal fragment of N-terminal
 CC haemagglutinating Moraxella catarrhalis outer membrane protein-106
 CC (OMP106). This sequence is used to design a probe and a 5' PCR primer.
 CC The OMP106 is used as a therapeutic and prophylactic vaccine against M.
 CC catarrhalis infections of mammals. It is used for diagnosis of bacterial
 CC infections and as reagents for clinical or medical diagnosis of M.
 CC catarrhalis infections and for scientific research on the properties of
 CC pathogenicity, virulence and infectivity of M. catarrhalis. It is also
 CC used as a probe to identify the presence of M. catarrhalis in biological
 CC specimens and to identify other bacteria that encode a polypeptide
 CC related to M. catarrhalis OMP106. OMP106-derived polypeptides are used as
 CC ligands to detect antibodies elicited in response to M. catarrhalis
 CC infections and also as immunogens for inducing M. catarrhalis-specific
 CC antibodies which are useful in immunoassays to detect M. catarrhalis in
 CC biological specimens. Cytotoxic antibodies are useful in passive
 CC immunisations against M. catarrhalis
 XX
 XX Sequence 24 AA;
 SQ

Query Match 57.9%; Score 121; DB 4; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EADGGKGGANARGDKSIAIGDIAQ 28
 |||||
 DB 1 EADGGKGGANARGDKSIAIGDIAQ 24

RESULT 8
 AAW04505
 ID AAW04505 standard; protein; 1992 AA.
 XX
 AC AAW04505;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-JAN-1997 (first entry)
 XX
 DE Moraxella 200 kDa outer membrane protein.
 XX
 XX Outer membrane protein; OMP; immunogen; vaccine; otitis media; diagnosis.
 KW Moraxella catarrhalis.
 OS
 XX WO9634960-A1.
 PN
 XX 07-NOV-1996.
 PD
 XX 29-APR-1996; 96WO-CA000264.
 PF
 XX 01-MAY-1995; 95US-00431718.
 PR 07-JUN-1995; 95US-00478370.
 PR 26-MAR-1996; 96US-00621944.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 XX Sasaki K, Harkness RE, Loosmore SM, Chong P, Klein MH;
 PI
 XX WPI; 1996-506162/50.
 DR N-PSDB; AAT38740.
 DR
 XX Moraxella outer membrane protein - useful as immunogen in protective
 PT vaccine and for diagnosis.
 PT
 XX Claim 14; Fig 6; 109pp; English.
 PS

XX An approx. 200 kDa outer membrane protein (AAW04505) can be isolated from
 CC Moraxella catarrhalis otitis media strain 4223 by electroelution, or
 CC expressed from a gene (see also AAF38740) obcd. from a strain 4223
 CC genomic library. Natural or recombinant outer membrane protein is useful
 CC as an immunogen to protect against infection by Moraxella, esp. M.
 CC catarrhalis. It can also be used to detect antibodies, esp. for
 CC differential diagnosis between bacteria that cause similar symptoms, and
 CC also useful as a carrier for other antigens and used to raise antitumour
 CC antibodies for conjugation to therapeutic agents. (Updated on 27-AUG-2003
 CC to correct OS field.)
 XX
 SQ Sequence 1992 AA;
 Query Match 43.1%; Score 90; DB 2; Length 1992;
 Best Local Similarity 52.3%; Pred. No. 0.024;
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
 QY 4 SEADGKGKANAR----GDKSTAIGDIAQALGQSQAIGDNKIV 43
 DB 1718 AKADGEAAVAIGRTQAGNQSTAIGDQAATGDSIAIGTGNV 1761
 RESULT 9
 AAB69137
 ID AAB69137 standard; protein; 1992 AA.
 XX
 AC AAB69137;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis M56 200kDa protein in pKS348 SEQ ID NO:13.
 XX
 KW Moraxella catarrhalis strain Q8; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; detection.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200107619-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-CA000870.
 XX
 PR 27-JUL-1999; 99US-00361619.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 XX
 DR WPI; 2001-159722/16.
 DR N-PSDB; AAF59106.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis.
 XX
 PS Claim 1; Fig 8A-V; 247pp; English.
 XX
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the M.
 CC catarrhalis M56 200kDa protein in pKS348, which is given in the

CC exemplification of the present invention
 XX
 SQ Sequence 1992 AA;
 Query Match 43.1%; Score 90; DB 4; Length 1992;
 Best Local Similarity 52.3%; Pred. No. 0.024;
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
 QY 4 SEADGKGKANAR----GDKSTAIGDIAQALGQSQAIGDNKIV 43
 DB 1718 AKADGEAAVAIGRTQAGNQSTAIGDQAATGDSIAIGTGNV 1761
 RESULT 10
 AAB69133
 ID AAB69133 standard; protein; 1992 AA.
 XX
 AC AAB69133;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE M. catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein SEQ ID NO:3.
 XX
 KW Moraxella catarrhalis strain 4223; major outer membrane protein;
 KW 200kDa outer membrane protein; antibacterial; immunogenic; infection;
 KW otitis media; detection.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200107619-A1.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-CA000870.
 XX
 PR 27-JUL-1999; 99US-00361619.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Sasaki K, Yang Y, Klein MH;
 XX
 DR WPI; 2001-159722/16.
 DR N-PSDB; AAF59100, AAF59101.
 XX
 PT New nucleic acid encoding Moraxella catarrhalis outer membrane protein,
 PT useful in protective vaccines and for diagnosis.
 XX
 PS Example 3; Fig 2A-W; 247pp; English.
 XX
 CC The present invention describes an isolated and purified nucleic acid (I)
 CC that encodes a 200 kDa outer membrane protein of Moraxella catarrhalis.
 CC The 200 kDa outer membrane protein (II) has antibacterial activity and
 CC can be used in vaccines. (II), and its truncated versions, are used as
 CC immunogenic compositions and vaccines to protect against M. catarrhalis
 CC infections, particularly otitis media in humans. (II) is also used as
 CC antigen in immunoassays for detecting specific antibodies (Ab), and to
 CC generate Ab. (I) are used for recombinant production of (II) and its
 CC fragments are used as probes for identifying/cloning 200 kDa protein
 CC genes from other strains, and for diagnostic detection of M. catarrhalis.
 CC (I) makes possible production of large amount of recombinant immunogens.
 CC Expression of truncated versions of (II) reduces toxicity of the protein
 CC towards the Escherichia coli host. The present sequence represents the M.
 CC catarrhalis strain 4223 lambdaEMBL3 clone 200kDa protein, which is used
 CC in the exemplification of the present invention
 XX
 SQ Sequence 1992 AA;
 Query Match 43.1%; Score 90; DB 4; Length 1992;
 Best Local Similarity 52.3%; Pred. No. 0.024;
 Matches 23; Conservative 5; Mismatches 12; Indels 4; Gaps 1;
 QY 4 SEADGKGKANAR----GDKSTAIGDIAQALGQSQAIGDNKIV 43
 DB 1718 AKADGEAAVAIGRTQAGNQSTAIGDQAATGDSIAIGTGNV 1761

for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX

Claim 25: SEO ID NO 62947: 1766pp: English.

XX

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence is encoded by one of the patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct_sequences

XX

Sequence 1946 AA;

Query Match	42.6%	Score 89:	DB 6:	Length 1946:
-------------	-------	-----------	-------	--------------

Best Local Similarity 52.3%: pred. No. 0.032:

DESC LOCAL SIMILARITY	92.5%	PID: NO: 0.6827
MATCHES	23: Conservative	4: Mismatches 13: Indels 4: Gaps

4 SEADGGKGGANAR----GDKSTATGDTAAAI.GSOSTAIGDNKIV 43

1672 AKADGEAAVAIGRQTQAGSQSIAIGDKAQAATGDSIAIGTCNVV 1715

Search completed: September 21, 2004. 23:39:48

Job time : 7.42676 secs


```
C;Genetics:
A;Gene: yopa
A;Genome: plasmid

Query Match      31.3%; Score 65.5; DB 2; Length 434;
Best Local Similarity 46.5%; Pred. No. 3.6;
Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGGANARGDK--SIAIGDIAQ-----ALGSQSIAIGDNKI 42
Db 94 GAGGLNARAKDPYSIALGATREAAKPAVAVGSGSIATGNSV 136

RESULT 3
C82672
surface-exposed outer membrane protein XF1516 [imported] - Xylella fastidiosa (strain 9a)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: C82672
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: C82672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1004 <SIM>
A;Cross-references: GB:AE003981; GB:AE003849; NID:g9106543; PIDN:AAF84325.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F.G.; Nunes, L.R.; Oliveira, M.A.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1516

Query Match      31.6%; Score 66; DB 2; Length 1004;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 15; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 9 GKGGANARGDKSIAIGDIAQALGSQSIAIG 38
Db 576 GSGVASAIGKDPSTATGASQAQVGDSSVALG 605

RESULT 4
S04534
invasin precursor - Yersinia pseudotuberculosis plasmid pTBI
C;Species: Yersinia pseudotuberculosis
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C;Accession: S04534; S04910
R;Roqvist, R.; Skurnik, M.; Wolf-Watz, H.
Nature 334, 522-525, 1988
A;Title: Increased virulence of Yersinia pseudotuberculosis by two independent mutations
A;Reference number: S04534; MUID:89302441; PMID:3043229
A;Accession: S04534
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <ROS>
A;Cross-references: EMBL:X12758; EMBL:X13883; NID:g48639; PIDN:CAA32088.1; PID:g48640
R;Skurnik, M.; Wolf-Watz, H.
Mol. Microbiol. 3, 517-529, 1989
A;Title: Analysis of the yopA gene encoding the YopI virulence determinants of Yersinia
A;Reference number: S04910; MUID:89343638; PMID:2761389
A;Accession: S04910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <SKU>
A;Cross-references: EMBL:X12758; EMBL:X13883; NID:g48639; PIDN:CAA32088.1; PID:g48640
```

```
C;Genetics:
A;Gene: yopa
A;Genome: plasmid

Query Match      31.3%; Score 65.5; DB 2; Length 434;
Best Local Similarity 46.5%; Pred. No. 3.6;
Matches 20; Conservative 3; Mismatches 11; Indels 9; Gaps 2;

QY 9 GKGGANARGDK--SIAIGDIAQ-----ALGSQSIAIGDNKI 42
Db 94 GAGGLNARAKDPYSIALGATREAAKPAVAVGSGSIATGNSV 136

RESULT 5
AH0110
probable surface protein (partial) YPO0902 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AH0110
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0110
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-658 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89747.1; PID:g15978974; GSPDB:GN00175
C;Genetics:
A;Gene: YPO0902

Query Match      31.3%; Score 65.5; DB 2; Length 658;
Best Local Similarity 39.0%; Pred. No. 5.5;
Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 2 GISEADGGKGGANARGDKSIAIGDIAQALGSQSIAIGDNKI 42
Db 480 GANSATCGAGSV-ASGNNSTAFSGKAKAANAALGANSV 519

RESULT 6
A82615
surface protein XF1981 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82615
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82615
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1190 <SIM>
A;Cross-references: GB:AE004017; GB:AE003849; NID:g9107083; PIDN:AAF84783.1; GSPDB:GN00
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F.G.; Nunes, L.R.; Oliveira, M.A.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328
A;Contents: annotation
```


R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J.C.; Dodson, R.; Gunn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: B81663
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <TET>
A;Cross-references: GB:AE002347; GB:AE002160; MID:g7190815; PIDN:AAF39597.1; PID:g719082
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0794
C;Superfamily: DNA-directed RNA polymerase alpha chain

Query Match 29.2%; Score 61; DB 2; Length 394;
Best Local Similarity 32.4%; Pred. No. 11;
Matches 12; Conservative 8; Mismatches 17; Indels 0; Gaps 0;

QY 3 ISEADGKGGANARGDKSIAIGDIAQAALGSGSIAIGD 39
Db 123 LQDCEGGRASQKLRTATISDASDLAAAGGQKATLGD 159

RESULT 12

H70589
Hypothetical glycine-rich protein Rv2853 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Mar-2003
C;Accession: H70589
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70589
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-615 <COL>
A;Cross-references: GB:295207; GB:ALJ23456; MID:g3261745; PIDN:CAB08453.1; PID:e315182;
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv2853
C;Superfamily: uncharacterized glycine-rich protein, PE motif containing

Query Match 29.2%; Score 61; DB 2; Length 615;
Best Local Similarity 41.5%; Pred. No. 17;
Matches 17; Conservative 4; Mismatches 18; Indels 2; Gaps 1;

QY 2 GISEADGKGGANARGDKSIAIG--DIAQAALGSGSIAIGDN 40
Db 408 GGAGNGGTGGAGPGGAGGAGGKADIANSLGDNATVTGGN 448

RESULT 13

F70868
Hypothetical glycine-rich protein Rv2487c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Mar-2003
C;Accession: F70868
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70868
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-694 <COL>

A;Cross-references: GB:AL021246; GB:AL123456; MID:g3261507; PIDN:CAA16064.1; PID:g2791515
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv2487c
C;Superfamily: uncharacterized glycine-rich protein, PE motif containing

Query Match 29.2%; Score 61; DB 2; Length 694;
Best Local Similarity 38.3%; Pred. No. 19;
Matches 18; Conservative 3; Mismatches 16; Indels 10; Gaps 2;

QY 2 GISEADGK---GGANARGDKSIAIGDIAQAAL-----GSGSIAIG 38
Db 292 GGAGGDDGTAAAGCGGGGNGDGGVAGDITASAFGGGNGSDGVAAG 338

RESULT 14

A49067
Transcription initiation factor IID chain p42 - fruit fly (Drosophila melanogaster)
N;Alternate names: TATA-binding protein (TBP)-associated factor TAFII40; TFIID complex
C;Species: Drosophila melanogaster
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Nov-2000
C;Accession: A49067; S42221
R;Goodrich, J.A.; Hoey, T.; Thut, C.J.; Admon, A.; Tjian, R. Cell 75, 519-530, 1993
A;Title: Drosophila TAF-II40 interacts with both a VP16 activation domain and the basal
A;Reference number: A49067; MUID:94037099; PMID:8221891
A;Accession: A49067
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-278 <GOO>
A;Cross-references: GB:L29540; MID:g463048; PID:g463049
A;Note: parts of this sequence were confirmed by peptide sequencing
R;Kokubo, T.; Gong, D.W.; Wootton, J.C.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y. Nature 367, 484-487, 1994
A;Title: Molecular cloning of Drosophila TFIID subunits.
A;Reference number: S42220; MUID:94150630; PMID:7545910
A;Accession: S42221
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-278 <KOK>
A;Cross-references: EMBL:U06458; MID:g458679; PIDN:AAC47347.1; PID:g458680
C;Genetics:
A;Gene: FlyBase:Taf40
A;Cross-references: FlyBase:FBgn0000617
C;Keywords: transcription initiation

Query Match 28.7%; Score 60; DB 2; Length 278;
Best Local Similarity 31.6%; Pred. No. 10;
Matches 12; Conservative 8; Mismatches 18; Indels 0; Gaps 0;

QY 1 IGISEADGKGGANARGDKSIAIGDIAQAALGSGSIAIG 38
Db 198 VGSGSGGGGGQEVKSESTGAGGDLKWEVDSDAAAVG 235

RESULT 15

T00797
Hypothetical protein At2g32710 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F24L7.15
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C;Accession: T00797; E84736
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, R. submitted to the EMBL Data Library, February 1998
A;Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A;Reference number: Z14204
A;Accession: T00797
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-327 <ROU>
A;Cross-references: EMBL:AC003974; MID:g2914688; PID:g2914702
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: E84736
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-327 <STO>
 A:Cross-references: GB:AE002093; NID:G2914702; PIDN:AC04492.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g32710; F24L7.15
 A:Map position: 2
 A:Introns: 193/2
 C:Superfamily: Arabidopsis thaliana hypothetical protein F24L7.15

Query Match 28.7%; Score 60; DB 2; Length 327;
 Best Local Similarity 33.3%; Pred. No. 12;
 Matches 12; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY 2 GISEADGGKGGANARGDKSIAIGDIAQALGQSIAI 37
 Db 12 GAGAGAGGGGGGGGGESSIALMDVVSPSSSLGV 47

Search completed: September 21, 2004, 23:44:38
 Job time : 3.42358 secs

